



Arnold Schwarzenegger
Governor

**MAINTAIN, ENHANCE AND IMPROVE
RELIABILITY OF
CALIFORNIA'S ELECTRIC SYSTEM
UNDER RESTRUCTURING**

APPENDIX - XV

California's Electricity Restructuring:
The Challenge to Providing Service
and Grid Reliability

Prepared For:

California Energy Commission
Public Interest Energy Research Program

Prepared By:

**Lawrence Berkeley
National Laboratory**

CERTS
CONSORTIUM FOR ELECTRIC RELIABILITY TECHNOLOGY SOLUTIONS

PIER FINAL PROJECT REPORT

MARCH 2006
CEC-500-2006-035-APXV
LBNL-58939



Prepared By:

Lawrence Berkeley National Laboratory
Consortium for Electric Reliability Technology
Solutions (CERTS)
Joseph H. Eto
Berkeley, CA
Contract No. 150-99-003

Prepared For:

California Energy Commission

Public Interest Energy Research (PIER) Program

Don Kondoleon,
Contract Manager

Mark Rawson, Bernard Treanton, Linda Kelly,
Ron Hoffman, Don Kondoleon
Project Managers

Mark Rawson
Program Area Team Lead

Laurie ten Hope
Office Manager,
ENERGY SYSTEMS INTEGRATION AND
ENVIRONMENTAL RESEARCH OFFICE

Martha Krebs, Ph. D.
Deputy Director
ENERGY RESEARCH AND DEVELOPMENT
DIVISION

B.B. Blevins
Executive Director

DISCLAIMER

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees or the State of California. The Energy Commission, the State of California, its employees, contractors and subcontractors make no warrant, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

California's Electricity Restructuring

The Challenge to Providing Service and Grid
Reliability

1007388

Final Report, December 2002

Cosponsors:

Public Interest Energy Research Program (PIER)
California Energy Commission
1516 Ninth Street
Sacramento, California 95814

PIER Project Manager:

D. Kondoleon

EPRI

3412 Hillview Avenue
Palo Alto, California 94304

EPRI Project Manager:

P. Hirsch

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

THIS DOCUMENT WAS PREPARED BY THE ORGANIZATION(S) NAMED BELOW AS AN ACCOUNT OF WORK SPONSORED OR COSPONSORED BY THE ELECTRIC POWER RESEARCH INSTITUTE, INC. (EPRI). NEITHER EPRI, ANY MEMBER OF EPRI, ANY COSPONSOR, THE ORGANIZATION(S) BELOW, NOR ANY PERSON ACTING ON BEHALF OF ANY OF THEM:

(A) MAKES ANY WARRANTY OR REPRESENTATION WHATSOEVER, EXPRESS OR IMPLIED, (I) WITH RESPECT TO THE USE OF ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT, INCLUDING MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, OR (II) THAT SUCH USE DOES NOT INFRINGE ON OR INTERFERE WITH PRIVATELY OWNED RIGHTS, INCLUDING ANY PARTY'S INTELLECTUAL PROPERTY, OR (III) THAT THIS DOCUMENT IS SUITABLE TO ANY PARTICULAR USER'S CIRCUMSTANCE; OR

(B) ASSUMES RESPONSIBILITY FOR ANY DAMAGES OR OTHER LIABILITY WHATSOEVER (INCLUDING ANY CONSEQUENTIAL DAMAGES, EVEN IF EPRI OR ANY EPRI REPRESENTATIVE HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES) RESULTING FROM YOUR SELECTION OR USE OF THIS DOCUMENT OR ANY INFORMATION, APPARATUS, METHOD, PROCESS, OR SIMILAR ITEM DISCLOSED IN THIS DOCUMENT.

ORGANIZATION(S) THAT PREPARED THIS DOCUMENT

**Mills College
Delft University of Technology**

CALIFORNIA ENERGY COMMISSION LEGAL NOTICE

THIS REPORT WAS PREPARED AS A RESULT OF WORK SPONSORED BY THE CALIFORNIA ENERGY COMMISSION (COMMISSION). IT DOES NOT NECESSARILY REPRESENT THE VIEWS OF THE COMMISSION, ITS EMPLOYEES, OR THE STATE OF CALIFORNIA. THE COMMISSION, THE STATE OF CALIFORNIA, ITS EMPLOYEES, CONTRACTORS, AND SUBCONTRACTORS MAKE NO WARRANTY, EXPRESS OR IMPLIED, AND ASSUME NO LEGAL LIABILITY FOR THE INFORMATION IN THIS REPORT; NOR DOES ANY PARTY REPRESENT THAT THE USE OF THIS INFORMATION WILL NOT INFRINGE UPON PRIVATELY OWNED RIGHTS. THIS REPORT HAS NOT BEEN APPROVED OR DISAPPROVED BY THE COMMISSION NOR HAS THE COMMISSION PASSED UPON THE ACCURACY OR ADEQUACY OF THIS INFORMATION IN THIS REPORT.

ORDERING INFORMATION

Requests for copies of this report should be directed to the EPRI Orders and Conferences, 1355 Willow Way, Suite 278, Concord, CA 94520, (800) 313-3774, press 2 or internally x5379, (925) 609-9169, (925) 609-1310 (fax). Copies of this report may also be downloaded from the California Energy Commission's website: <http://www.energy.ca.gov/research>

Electric Power Research Institute and EPRI are registered service marks of the Electric Power Research Institute, Inc. EPRI. ELECTRIFY THE WORLD is a service mark of the Electric Power Research Institute, Inc.

Copyright © 2002 Electric Power Research Institute, Inc. All rights reserved.

CITATIONS

This report was prepared by

Mills College
5000 MacArthur Boulevard
Oakland, CA 94612

Delft University of Technology
P.O. Box 5015, 2600GA
Delft, The Netherlands

Principal Investigators

E. Roe
M. van Eeten
P. Schulman
M. de Bruijne

This report was prepared for

EPRI
3412 Hillview Avenue
Palo Alto, California 94304

and

Public Interest Energy Research Program (PIER)
California Energy Commission
1516 Ninth Street
Sacramento, CA 95814

This report describes research jointly sponsored by EPRI and the California Energy Commission.

The report is a corporate document that should be cited in the literature in the following manner:

California's Electricity Restructuring: The Challenge to Providing Service and Grid Reliability,
EPRI, Palo Alto, CA, California Energy Commission, Sacramento, CA: 2002. 1007388.

REPORT SUMMARY

For all the attention devoted to the California energy crisis, the challenge restructuring posed to the reliability of the electricity generation, transmission, and distribution system has largely been overlooked. Reliability is often taken for granted even though an uninterrupted power supply, the fundamental precondition of the energy market, depends on the determined, real-time efforts of professionals in the control rooms where load and generation are balanced. This report looks at restructuring and its continuing aftermath from a reliability perspective.

Background

The untold story of the California electricity crisis is what it looked like from within the control rooms of those who kept the lights on for the California public and prevented system failure on the western grid. To understand how restructuring permanently changed the functional meaning of grid and service reliability, it is necessary to understand what went on during the crisis in the control rooms of the California Independent System Operator (ISO) and the control rooms of generators and distribution utilities. This report examines how the control rooms actually kept the lights on virtually all the time during the crisis, albeit "just-in-time" and with not much time to spare. It also looks to the implications of this experience on the question of how private-sector generators, public-sector regulators, and not-for-profit grid managers can provide electricity reliably in the future despite their competing goals and interests.

Objectives

To analyze California's electricity restructuring and crisis from the perspective of grid and service reliability under persisting real-time performance conditions; to examine the continuing institutional design implications of the changing nature of electricity reliability in volatile markets.

Approach

As part of a project on how critical infrastructures ensure high service reliability under rapidly changing conditions, researchers studied how grid and service reliability were maintained in real time during California's electricity crisis. They based their analysis on 60 interviews with ISO control room personnel and representatives of utilities, energy trading companies, and regulatory agencies, among others. Recognizing the critical role of reliability professionals in maintaining the grid in real time, the team made recommendations focused on developing this emerging profession through improved training, communication, and public recognition.

Results

Deregulated markets make it much more difficult and expensive to coordinate interdependency among the many market participants. Energy trading thrives on price volatility, but that volatility is the enemy of real-time reliability when balancing load and generation in the control rooms. Indeed, for many grid operators and managers, the all-absorbing preoccupation with real-time operations is what defined the electricity crisis. The proportions of time spent between real-time operations and longer-term management, planning, and investment have permanently changed because of the California electricity restructuring and the crisis it induced. The new emphasis on real-time operations has already had consequences. For example, the ISO staff has been greatly increased. Adjusting to the new real-time burden and balancing it with other strategies for grid management have become a permanent feature of the California power system.

Any proposed improvement in the California electricity system should be examined to see if its implementation will increase system volatility, decrease the available options to balance load and generation in response to that volatility, or interfere with the adaptability of those responsible to maintain and ensure grid and service reliability. A proposal must pass the reliability-matters test. However, improving reliability in the system may depend more upon fostering the professionalism and expertise of the people responsible for real-time operations than on any change in organizational design.

EPRI Perspective

Reliable electricity has become a critical infrastructure; American society is simply unwilling to trade off reliability against electricity's other features. Markets that operate within a context that takes grid and service reliability seriously are the only reason why we could ever expect grid and service reliability to be enhanced through those markets. The generation assets sold off under electricity restructuring were worth so much precisely because they were tacitly backed up by a governmental willingness to ensure that reliability of electricity would not suffer as a consequence. The experiment of not taking reliability seriously has led Californians and others to pay a very high price to make good that guarantee. Critical infrastructures require highly competent, always-on management, whether or not the "prices are right."

Many policy proposals are currently being made to change the electricity system, but recognizing and supporting the professionals responsible for reliability management may be more important than particular design issues. The real story of the California electricity crisis is not about structures, but about the skills that matter under changing performance conditions in real time. One way of improving the recruitment and training of key personnel could be the creation of a state-wide Electricity Reliability Network of reliability specialists, practitioners and regulators who could speak on behalf of reliability for the California electricity system as a whole.

Keywords

Deregulation

Reliability

Independent Service Operator (ISO)

ACKNOWLEDGMENTS

The authors of this report are Emery Roe, Michel van Eeten, Paul Schulman and Mark de Bruijne. Our report is based on original research gathered prior to our working with EPRI and the CEC and thus covers only material primarily related to their interest in real-time reliability issues affecting California's electricity system.

A report such as this incurs many debts along the way. Special thanks go to David Hawkins without whose help and encouragement it would have been very difficult to write about this topic. We were extremely fortunate to have access to the CAISO control room through the good faith of Jim Detmers and Jim McIntosh and the forbearance of the control operators we observed and interviewed there.

Two of our interviewees asked to remain anonymous. We would like to thank them as well as the others who also gave generously of their time: Ziad Alaywan, Massoud Amin, Ali Amirali, Steve Auradou, Diane J. Barney, Barry Bauer, Larry Bellknap, Tracy Bibb, Boyd Busher, Ron Calvert, Russ Calvery, Karen Cardwell, Kevin Coffee, David Delparte, Terry P. Dennis, Jim Detmers, Patrick Dorinson, Tami Elliot, Joseph H. Eto, Paul Feeley, Mike Fling, Kellan Fluckiger, Mark Franks, Lee S. Friedman, Steve Gillespie, Mary Glas, Hector Gonzalez, Barbara Hale, Duane Haushild, Dave Hawkins, Christine Henry, Lauri Jones, Stan K. Kataoka, Jamie J. Labban, Don LaDue, Stephen T. Lee, Lawrence Lingbloom, Jim McIntosh, Kim Malcolm, Stephen A. MacCool, Jeffrey C. Miller, Marty Moran, Dale Murdock, Randy Nunnink, Paul Olson, Shmuel S. Oren, Brian S. Rahman, Jason Regan, Guy Ridgely, Mark Rothleder, Mike Starr, Steven Stoft, Ed Street, Robert Stuart, Robert E. Sullivan, Bob Therkelsen, Brian Thurston, Tim Van Blaricom, Greg Van Pelt, Danielle Vockeroth Smith, Keith A. Wheeler, Terry Winter and Frank Wolak. They bear no responsibility for the way we have interpreted their interviews.

We also owe a great deal to the high reliability theory crew. Todd LaPorte took time to observe the CAISO control room and has served as a steady sounding board for our evolving ideas and framework. An earlier lunch with Karlene Roberts was encouraging and Gene Rochlin has always had something important to say about the work. We have given presentations to conferences on our findings, including the 5th International Conference on Technology, Policy and Innovation, the 24th Annual Research Conference of the Association for Public Policy Analysis and Management, and a CPUC/GSPP conference on the California electricity restructuring and crisis. Peter Hirsch, Don Kondolean and Joe Eto not only saw to the funding of the report but made helpful comments on the draft version. The standard disclaimer also applies here, much as we would like to displace error onto others and get credit only for the good parts.

EXECUTIVE SUMMARY

Synopsis of Argument and Findings

This report presents the findings on the reliability of electricity provision under performance conditions arising out of California's electricity restructuring of the late 1990s.

The research was undertaken as part of the Networked Reliability Project, a joint initiative involving faculty and researchers from Mills College in Oakland and Delft University of Technology in the Netherlands. We undertook the bulk of investigations in the form of hour or longer interviews between April and December, 2001. Sixty interviewees were identified and interviewed: thirty-three in and around the Folsom control room of the California Independent System Operator (ISO); eight at PG&E (e.g., in and around their Transmission Operations center and the Operations Engineering units); five with a large generation supplier (a senior generation official and control room operators in one of its large California plants) and a private market energy trading dot com; and fourteen others located in the Governor's Office, California Public Utilities Commission (CPUC), California Energy Commission (CEC), Electric Power Research Institute (EPRI), Lawrence Berkeley National Laboratory (LBNL), University of California, Berkeley (UCB), and Stanford. More details on the research methodology follow in the Executive Summary's Annex.

The specific objective of our research has been to provide a framework for analyzing, discussing the principal features of, and detailing the major factors pushing and pulling California's restructured electricity network into the management of reliable electricity under real-time performance conditions, i.e., those factors that encourage balancing load and generation in this hour or the next hour ahead.

The specific objective follows from the overarching research question of the Networked Reliability Project: How can critical infrastructures, such as electricity, water supplies, financial services, and telecommunications, ensure and maintain high service reliability under deregulated or increasingly fragmented conditions? Such developments make provision of peak load services safely and continuously through time more and more dependent on a widening set of networked organizations operating within the infrastructures. In the California restructuring case, the research question became: How does the network of deregulated generators, transmission managers and distribution utilities, many with competing if not conflicting goals and objectives, achieve highly reliable electricity service over time; and, more especially, how does that network ensure high reliability provision in real time, when load and generation are balanced only in the current or next hour rather than in the day ahead or hour ahead markets created through restructuring?

From a reliability perspective, the California electricity system is described in terms of three components: the reliability task environment (*RTE*) and within it, the regulatory reliability network (*RRN*) and the high reliability network (*HRN*). Together, the *RTE*, *RRN* and *HRN* constitute what we term the *California electricity system*.

The report is almost exclusively focused on the high reliability network created by California's electricity restructuring (see Executive Summary's Annex for our report audiences). Simply put, the *HRN* organizes for electrical service provision and delivery and includes the organizations and units that have direct operational responsibilities for this provision and delivery. In the California electricity system at the time of our study, the *HRN* consisted of the control rooms and support staff of the ISO, the distribution utilities, the private generators along with scheduling coordinators (market traders and California Energy Resources Scheduling [*CERS*] division) and the adjacent control areas.

The *RRN* sets the mandates and criteria for high reliability, i.e., it establishes and/or enforces standards that define reliability for units in the *HRN*. The standards include formal regulations under a government mandate and regulations developed and enforced by system participants themselves, such as *WSCC* (Western Systems Coordinating Council) standards. The *RRN* does not organize the provision of services directly. In the California electricity system, the *RRN* consists of the *CPUC*, Federal Energy Regulatory Commission (*FERC*), *WSCC*, North American Electricity Reliability Council (*NERC*), *CEC* and the Electricity Oversight Board (*EOB*), among others. While these agencies and units have reliability of electrical service as a mandate, they often also have other mandates to reconcile, e.g., cost to electricity consumer.

The *RTE* sets the context for the reliability-related tasks of the *RRN* and *HRN*. The reliability task environment includes customers, voters, businesses, elected officials and the public. The customer has demands, expectations, and even certain contractual rights to receive service in a reliable fashion. A widespread notion in California is that electricity is not just any other commodity and that access to cheap electricity is an entitlement, if not right. Clearly, the *RTE*, *RRN* and *HRN* are connected, as when the public's pervasive notion of electricity as a right influenced much of the intervention of the Governor's Office to have *CERS* fund the ISO's provision of electricity during the California electricity crisis.

This report's focus is on the implications of real-time operations for the provision of electricity on the reliability of that electricity. The following is a summary of our *HRN*-related findings.

The answer to our core question—How does the *HRN* maintain reliable electricity in real time?—is this: It balances load and generation in real time by developing and maintaining a repertoire of responses and options in the face of unpredictable or uncontrollable system volatility, where the “system” in question is the California electricity system as defined above.

More especially, our report's framework focuses on the match between, on one hand, the options and strategies of the *HRN* to achieve its reliability requirement (e.g., balancing load and generation, staying within thermal limits set for key paths) and, on the other hand, the unpredictable or uncontrollable threats to fulfilling the *HRN* reliability requirement. A match results from having at least one option sufficient to meet the requirement under given conditions.

At any point, there is the possibility of a mismatch between the system variables that must be managed to achieve the reliability requirement and the options and strategies available for managing those variables. As the following chapters underscore, that match is not automatic and requires management—in this case high reliability management within the HRN.

Within the California electricity HRN, the mandate to balance load and generation, the core reliability task of the HRN, is located in the focal organization, CAISO. It is the only organization that simultaneously has the two reliability mandates: keeping the electricity flow always on and reliably protecting the grid from islanding and meltdown. Meeting the dual reliability mandate involves managing the options and strategies that coordinate actions of the independent generators, energy traders and utilities in the HRN. As the focal organization, the options the ISO deploys are HRN-based or HRN-wide options, e.g., outage coordination is the responsibility of the ISO, but involves the other partners in the HRN.

Consequently, the ISO management can be categorized in terms of the variety of HRN-based options it, the ISO, has available (high or low) and the volatility of the California electricity system (high or low), resulting in four performance conditions and modes as set out in Figure ES-1.

		System Volatility	
		<i>High</i>	<i>Low</i>
Network Option Variety	<i>High</i>	Just-in-time performance	Just-in-case performance
	<i>Low</i>	Just-for-now performance	Just-this-way performance

Figure ES-1
Performance Conditions for the Focal Organization, the ISO

Volatility is the degree to which the focal organization, the ISO, faces uncontrollable changes or unpredictable conditions that threaten the grid and service reliability of electricity supply, i.e., that threaten the task of balancing load and generation. Some days are of low volatility. A clear example of high volatility are those days where a large part of the forecasted load had not been scheduled through the day-ahead desk, which means for the ISO actual flows are unpredictable and congestion will have to be dealt with at the last minute. Volatility refers to any system-related instabilities, not to price movements alone.

Options variety is the amount of HRN resources, including strategies, available to the ISO to respond to events in the system in order to keep load and generation balanced at any specific point in time. It can be approximated with conventional engineering parameters, including available operating reserves and other generation capacity and available transmission capacity. High option variety means, for instance, that the grid has more than the required regulatory resources available (there are high or wide margins), low options means the resources are below

requirements and, ultimately, that very few resources are left (low or tight margins). These two dimensions together set the conditions under which the ISO has to operate and demand different performance modes for ensuring reliability (i.e., for ensuring the balancing of load and generation): "just-in-case," "just-in-time," "just-for-now," and "just-this-way."

- *"Just-in-case" performance mode for balancing load and generation:* When options are high and volatility low, "just-in-case" performance is dominant because of high redundancy. Reserves available to the ISO are large, excess plant capacity exists at the generator level, and the distribution lines are working with ample backups, all much as forecasted with little or no volatility (again, unpredictability and/or uncontrollability). More formally, redundancy is a state where the number of different but effective options to balance load and generation is high relative to the market and technology requirements for this balance. There are, in brief, a number of different options and strategies to achieve the same balance. The state of high redundancy is best summed up as one of maximum equifinality, i.e., there are very many means to meet the reliability requirement, "just in case" they are needed. Yet, as we shall see, it is important not to confuse this mode with the pre-restructuring condition where integrated utilities had "high reserves."
- *"Just-in-time" performance mode for balancing load and generation:* When options and volatility are both high, "just-in-time" performance is dominant. Option variety to maintain load and generation remains high, but so is the volatility in system variables. High volatility may be in the form of underscheduling or rapid price fluctuations leading to unexpected strategic behavior by market parties, while higher grid volatility may be in the form of contingencies such as sagging transmission lines during unexpected hot weather. This performance condition demands real-time flexibility, that is, the ability to utilize and develop different options and strategies quickly in order to balance load and generation. Operators in the control room are in constant communication with each other and others in the HRN, options are reviewed and updated continually, and informal communications are much more frequent. Flexibility in real time is the state where the operators are so focused on meeting the reliability requirement and the options to do so that more often than not they "customize" the match between them, i.e., the options are just enough and "just-in-time."

More formally, the state of real-time flexibility is best summed up as adaptive equifinality: There are effective alternative options, which are developed or assembled as required to meet the reliability requirement. Substitutability of options and strategies is high for "just-in-time" performance, where the increased volatile network behavior is matched by the flexibility in options and strategies for keeping performance within formal reliability tolerances and bandwidths.

- *"Just-for-now" performance for balancing load and generation:* When option variety is low but volatility is high, "just-for-now" performance is dominant. It is the most unstable performance mode of the four and the one control room operators want most to avoid or get out of as soon as possible. Options to maintain load and generation have become visibly fewer and increasingly insufficient to what is needed in order to balance load and generation. This state could result from various reasons related to the behavior of the electricity system. Unexpected outages can occur, load may increase to the physical limits of transmission capacity; and the use of some options can preclude or exhaust other options, e.g., using stored hydro capacity now rather than later. Under these performance conditions,

unpredictability or uncontrollability has increased (i.e., volatility has increased), with variety of effective options and strategies diminished or less available. For example, a Stage 1 or 2 emergency has been declared by the ISO and a senior ISO official goes outside of official channels and calls his counterpart at a private generator, who agrees to keep the unit online, “just for now.”

More formally, "just-for-now" performance is a state best summed up as one of maximum potential for “deviance amplification:” Even small deviations in elements of the market, technology or other factors in the system can ramify widely throughout the system. Marginal changes can have maximum impact in threatening the reliability requirement, i.e., the loss of a low-megawatt generator can tip the system into blackouts. From the standpoint of reliability, this state is untenable over an extended period of time. Here people have no delusions that they are in full control. They understand how vulnerable the network is, how limited the options are and precarious the balance, they are keeping communications lines open to monitor the state of the network, and they are busily engaged in developing options and strategies to move out of this state. They are not panicking and, indeed, by prior design, they still retain the crucial option to reconfigure the electricity system itself, by declaring a Stage 3.

- *"Just-this-way" performance for balancing load and generation:* When options variety and volatility are both low, "just-this-way" performance is dominant. This performance condition occurs in the California electricity system as a short-term “emergency” solution. In an electricity crisis, the option is to tamp down volatility directly with the hammer of crisis controls and forced network reconfigurations. The ultimate instrument of crisis management strategy is acknowledged to be the Stage 3 declaration, which requires interruption of firm load in order to bring back the balance of load and generation from the brink of "just-for-now" performance.

More formally, "just-this-way" performance is a state best summed up as one of zero equifinality: Whatever flexibility could be squeezed through the remaining option and strategies is forgone on behalf of maximum control of a single system variable, in this case load. The Stage 3 declaration has become both a necessary and sufficient condition for balancing load and generation, in this case reducing load directly. This contrasts significantly with the other three performance conditions. There the options and strategies are sufficient, without being necessary. Many ways exist to skin the cat. Only in "just-this-way" performance is the “remaining option” both sufficient and necessary. You are left with only one way, or no way.

For the purposes of the report, we term "just-in-time" and "just-for-now" performance modes of balancing load and generation under high system volatility (i.e., the left side of the Figure ES-1 typology) to be, “real-time reliability.” The differences between the performance modes, along with others discussed more fully in the report, are summarized in Table ES-1.

When we returned to the ISO in mid-2002 to present the results of our research—this being a year after our control room observations during the peak of the 2001 electricity crisis—we were informed that well over 85% of control room activity was still in the real-time reliability performance modes of "just-in-time" and "just-for-now." Granted price volatility was no longer

the problem it had been in April 2001, but now it was a set of problems of “something else” arising out of the electricity restructuring.

Our research findings and subsequent presentations to staff at the ISO and CPUC lead us to believe that “high system volatility” is here to stay for the foreseeable future as a result of California’s restructuring exercise. The reasoning for this is detailed in the following chapters and accounts for the length of our report. The system volatility will look different from time to time, and there will be occasions when things quiet down, but high system volatility or instability has become a permanent feature of the HRN, we believe. Some of the more important features pushing and pulling control rooms into real-time reliability conditions are summarized in Table ES-2.

Table ES-1
Selected Features of Four Performance Modes For Reliability

	Performance mode			
	Just-in-case	Just-in-time	Just-for-now	Just-this-way
Volatility	low	high	high	low
Option variety	high	high	low	low
Principal Feature	high redundancy	real-time flexibility	max. potential for amplified deviance	command & control
Operational risks	risk of inattention & complacency	risk of misjudgment because of time & system constraints	risk of depleted options & lack of maneuverability (most untenable mode)	risk of control failure over what needs to be controlled
Information strategy	vigilant watchfulness	keeping the bubble	localized firefighting	compliance monitoring
Lateral communication	little lateral communication during routine operations	rich, lateral communication for complex system operations in real-time	lateral communication around focused issues and events	little lateral communication, during fixed protocol (close to command & control)
Rules and procedures	performing according to wide-ranging established rules and procedures	performing in & outside analysis; many situations not covered by procedures	performing reactively, waiting for things to happen, "I'm all tapped out"	performing to very specific set of detailed procedures
Orientation toward Area Control Error	having control	keeping control	losing control	forcing command & control

Table ES-2
Selected Push & Pull Factors for HRN Real-Time Reliability

-
- Real-time is an answer to persistent network incompleteness
 - Positive interdependencies are most evident & likely to be acted on in real-time
 - Real-time is informal, non-routine, & flexible
 - Real time allows for larger process variance relative to output variance around the balance of load & generation
 - Improvisation is maximized and rewarded
 - Service reliability is non-fungible except when risking grid reliability in real time
 - Shared professional norms, information & accountability substitute for the culture of reliability in real-time
 - Flexible authority patterns and teams are locus of highly reliable behavior
-

These factors as well as others are explained more fully in the report; some factors, as the “incomplete network design” and “non-fungibility” are so significant that we have devoted individual chapters to them. The Annex gives a summary of the report’s chapters.

What does all of this mean for decision makers, particularly those at the CEC, LBNL and EPRI who have funded our research?

First, let us summarize where we are at this point. How does the HRN actually produce, transmit and deliver reliable electricity within the California electricity system under varying performance conditions? Answer: By having the focal network organization—the ISO—able to (1) maintain the balance of load & generation within any one of the four performance modes of options variety and system volatility; (2) adapt to external-generated shifts in system variables so as to maintain the balance of load and generation; (3) move out of “just-for-now” performance mode as soon as possible into “just-in-time,” “just-this-way,” or “just-in-case” performance mode; and (4) develop longer-term strategies to increase network options and/or reduce system volatility.

In short, reliability (both service and grid) results when performance is sustained across all four HRN conditions. Also, cross-performance adaptability—being able to shift on the fly—is as important to high reliability as balancing load and generation in any one performance mode.

If we are right in the above, then we can propose a *reliability-matters test* that any proposed “improvement” in the California electricity system should pass before it is adopted: Does it reduce system volatility, increase network options, enhance ISO cross-performance adaptability, and/or lead to effective longer term strategies to do the same? We give examples of putative “system improvements” that would increase volatility, reduce options and limit adaptation, now and in the future.

Finally, if service and grid reliability do matter, then we should be paying much more attention to those professionals who treat them seriously. We call this group of people, reliability professionals. They may be economists, engineers, or those who worked their way up into the control rooms from field crews. They may have degrees or no degrees, but they all have the knowledge bases, experience and skills to treat real-time reliability as seriously as it deserves, given the electricity is a major critical infrastructure of society. They are found throughout the high reliability network, sometimes even in the regulatory agencies (RRN). We offer a number of recommendations, including undertaking what we call “reliability analysis,” to heighten the prominence and importance of these professionals in the restructured electricity system that California now has.

In sum, we offer no administrative reform, policy initiative, sector restructuring, technological fix or institutional redesign to improve the California electricity system, largely because the proposals we see out there on the radar screen would not pass the reliability-matters test. Instead, we believe our recommendations for recognizing and enhancing the analytic capacity, career development, and recruitment of the reliability professionals go to the heart of safeguarding our electricity critical infrastructure.

Annex

Research Methods. Following the practices of policy analysis and social science, we relied on multiple methods, documents and key informants to identify and crosscheck answers to our research questions. A two-phased study for our research was adopted. In early 1999, we reviewed the literature on deregulation of the energy sector, with special reference to California’s electricity restructuring. During this initial phase, we identified CAISO as a key focal organization for our primary research. We approached key officials there and received permission to interview staff in and around the ISO main control room. Key informant interviews proceeded by the snowballing technique, in which new interviewees were identified as “people we should talk to” by previous interviewees, until we reached a point where new interviewees were mentioning the same or similar problems and issues we had heard in previous interviews. An interview questionnaire was used throughout (allowing open-ended responses and follow up), with face-to-face interviews typically lasting an hour or more. Also a major portion of this report is based on direct observation, simply watching what individuals did in their jobs through many shifts in performance conditions.

Report Audiences. The report has been written for staff at EPRI, LBNL and CEC. It should also be of interest to officials and staff in the ISO, distribution utilities and private generators that operate within the California electricity system. Because of its topicality and subject matter, the report will also appeal to a wider audiences, including: economists working in deregulation, policy analysts and public managers involved in institutional design of governance structures for large technical systems, engineers mandated to design and operate critical infrastructures, organization theorists analyzing technological accidents and organizational reliability, social scientists studying major technology transformations, and planners engaged in long-term energy initiatives.

Report Chapters. The report has ten chapters. Chapter 1 provides the background and roadmap to the report. Chapter 2 sets out the overall empirical context and conceptual framework for the report. We discuss the major theoretical and disciplinary explanations for why service reliability should be difficult to achieve under restructured and crisis conditions. Each explanation has its shortcomings and we present our own conceptual framework that provides, we believe, a more effective explanation of why service and grid reliability was much better than the major theories and disciplines would predict. Our framework is organized around the dimensions of system volatility and network options to respond to that volatility, and identifies the four primary performance modes mentioned earlier to achieve reliable electricity through balancing load and generation.

Chapters 3-9 examine the important features and reliability challenges to the California electricity system identified through our framework, interviews, observations and research. The design features are important both for maintaining real-time reliability (again, "just-in-time" and "just-for-now" performance under conditions of high system volatility) and for constituting challenges to the improvement of that reliability.

Chapter 3, "The wraparound of market and technology," reveals the substantial organizational infrastructure of ISO programmers, engineers, lawyers, economists, and others needed to enable the electricity market to function and connect its outcomes to the physical properties of the grid, including buffering the ISO control room operators against wider system volatility. Chapter 4, "Electricity as an always-on service," highlights the reliability consequences of the service's "always-on" requirement. Even though the "culture of reliability" no longer dominates in the network-wide provision of electricity (as it did in key parts of the earlier integrated utilities), we did find official and unofficial strategies of shared knowledge, timeliness and accountability for actions and their consequences as a result of the "always-on" feature of management.

Chapter 5, "Persistently incomplete design," questions the prevailing view that the electricity crisis was the result of poorly designed institutions and that once the missing elements are provided (e.g., more generation or retail deregulation), the California electricity system will become sufficiently stable and predictable. Our research reveals two major mechanisms that underlie and maintain an enduring incompleteness of policy and institutional design in the restructured California electricity system: both the overall complexity and the internal dynamics of any such designed network. The conclusion of the chapter is that design incompleteness is permanent and that improved mechanisms to cope with it are essential to ensuring reliability under enduring real-time performance constraints.

Chapters 6 through 9 describe four features which are increasingly important for the ISO in "pulling things together hour by hour up until the last minute" in order to ensure real-time reliability of electricity. First, in order to have electricity be always on it is treated by the ISO—more specifically its control room operators—as non-fungible in real time. Chapter 6, "Non-fungibility of reliability in real time," describes how the demand for both service and grid reliability makes electricity not "just another commodity"—which has been a basic assumption behind deregulation. In Chapter 6 we show that grid reliability is non-fungible to such an extent that the state was unwilling to trade off service reliability (keeping the lights on) except only in those few instances where shedding load and controlled blackouts were needed to preserve grid

reliability. The non-fungibility of grid reliability in real time is a primary reason why California electricity was provided as reliably as it was during the crisis, even under market meltdown and controlled blackouts.

Chapter 7, “Markets and their fallbacks,” uses our interviews and observations to show how the ISO control room and its wraparound actually supplemented markets and market transactions at the height of the crisis and also undertook other interventions that compensated for market problems and failures. When the day-ahead, hour-ahead and real-time market desks were breaking down *ad seriatim*, the ISO managed to fill the coordination gap left behind in order to maintain grid and service reliability with a variety of different market fallback mechanisms. Chapter 8, “(Re)defining reliability,” addresses the feature whereby definitions, standards and criteria of the reliability of the system are changing. Our findings run counter to the belief that a breaching of official standards looks to be *prime facie* a decline in reliability. An important part of this redefinition, we argue, is adapting the standards to conditions of high volatility—in the words of one interviewee, to give the ISO a “bar that they can jump.”

Chapter 9, “The push and pull to real time,” summarizes sixteen factors at work in pushing and pulling the balancing of load and generation together at the last minute. While the original design for the restructured electricity system envisioned that real-time operations would deal with the last few percent of electricity transactions, a large portion of overall electricity demand was provided hour by hour or at the last moment during the crisis. Many insiders, who feel pushed into real time as the last resort to prevent grid collapse, have considered this a failure in reliability. Our report underscores not just the push but also the pull factors as well driving the California electricity system into real time, hour-by-hour, if not last-minute provision of services, factors that will persist notwithstanding new generation and new market structures to come on line in the future.

Chapter 10 recommendations offer no silver bullet for improving the California electricity system. We believe that many, if not most, current proposals would not pass the reliability-matters test. In the process, we give examples of “reliability analyses” of proposals for Stage 3 declarations, real-time residential metering, and homeland security that show how so-called improvements could actually worsen grid and service reliability. Our own recommendation—enhancing the analytic capacity, career development, and recruitment of the reliability professionals who ensure the system’s reliability—is, we believe, at the heart of truly safeguarding and improving our electricity critical infrastructure.

CONTENTS

- 1 INTRODUCTION AND ROADMAP 1-1**
 - Introduction 1-1
 - The California Electricity Crisis 1-3
 - Study Project, Research Question and Methods 1-6
 - Roadmap 1-7

- 2 CONTEXT AND FRAMEWORK 2-1**
 - Context 2-1
 - Framework 2-5
 - The High Reliability Network 2-7
 - How Does the HRN Produce, Transmit and Deliver Reliable Electricity within the California Electricity System Under Varying Conditions? 2-9
 - “Just-in-Case” Performance, Redundancy and Maximum Equifinality 2-11
 - “Just-in-Time” Performance, Real-Time Flexibility and Adaptive Equifinality 2-12
 - “Just-for-Now” Performance and Maximum Potential for Deviance Amplification 2-12
 - “Just-this-Way” Performance, Crisis Management and Zero Equifinality 2-13
 - Transitions Between the States of Performance 2-17

- 3 THE WRAPAROUND OF MARKET AND TECHNOLOGY 3-1**
 - Introduction 3-1
 - The Wraparound in Action 3-3
 - Four Qualifications about the Wraparound 3-6
 - The Wraparound in “Just-in-Time” Performance 3-8
 - Conclusion 3-10

- 4 ELECTRICITY AS AN ALWAYS-ON SERVICE 4-1**
 - Introduction 4-1
 - What Does “Always-On” Mean? 4-1

Always-On Management as Cross-Performance Adaptability	4-3
Feedback on Performance	4-5
Keeping the Network Honest and Accountable	4-9
Factors Important for Trust and Accountability	4-11
The Importance of Shared Communications	4-14
Conclusion	4-17
5 PERSISTENTLY INCOMPLETE DESIGN	5-1
The Missing Pieces	5-1
Incompleteness in HRN, RRN, and RTE	5-2
“We Are In a Transition Right Now...”	5-5
Persisting Incompleteness	5-6
Negative and Positive Implications for Reliability	5-9
Conclusion	5-13
6 NON-FUNGIBILITY IN REAL TIME.....	6-1
Not Just a Commodity	6-1
From Cost-of-Service Reliability to Reliability-at-Any-Price.....	6-4
Why Didn't Anyone Think About Whether Restructuring Would Lead to Higher Prices?	6-5
Non-Fungibility in Real Time	6-6
Integrating Non-Fungibility in the Institutional Design	6-8
7 MARKETS AND THEIR FALLBACKS	7-1
And Then One Day, the Markets Were Gone	7-2
Markets on Life Support	7-4
Rebuilding the Links: Fallbacks for the Market Design	7-6
Informal Networks	7-7
Contingent Command and Control	7-10
Organizational Adaptation	7-12
Conclusion	7-13
8 (RE)DEFINING RELIABILITY	8-1
Adapting Reliability Standards to New Conditions	8-1
Redefinition in the RTE.....	8-1
Redefinition in the RRN.....	8-2

Redefinition in the HRN	8-4
Adapting Reliability Standards to Real-Time Conditions	8-4
Redefinition: What is Reliability in the HRN?	8-6
Again, Has Reliability Gone Down?	8-11
Implications	8-14
9 THE PUSH AND PULL TO REAL TIME	9-1
Introduction	9-1
Push to Real Time	9-1
Pull to Real Time	9-3
1. Real Time is an Answer to Persisting Network Incompleteness	9-4
2. Positive Interdependencies are Most Visible in Real Time	9-5
3. Real Time Remains Informal, Non-Routine and Relational	9-5
4. The Need for Small Output Variance (Load and Generation Must Be Balanced) is Adapted to by Real-Time Multiple Means with Large Process Variance (Equifinality)	9-7
5. Real Time Legitimizes and Accommodates Redefining Reliability Criteria	9-9
6. Reliability is Non-Fungible in Real Time	9-10
7. Real-Time Operations Justify Improvisation and Experimentation	9-11
8. HRN Tolerance of What Would Otherwise be Considered Low Performance is Greatest in Real Time	9-13
9. While Often Negative, Strategic Behavior Can Have Positive Effects on Real- Time Reliability	9-14
10. Market and Deregulation Incentives Exist for Ensuring Real-Time Reliability	9-15
11. Pulling New Participants into the HRN Becomes an Option for Increasing Real- Time Reliability	9-16
12. Shared Professional Norms, Competence, Information and Accountability Substitute for the Culture of Reliability in Real Time	9-17
13. Reliance on Real-Time Information and Software Leads to Reliability in Real Time	9-21
14. Flexible Authority Patterns and Teams Become the Locus of Reliable Behavior	9-22
15. HRN Relations with the RRN and RTE Seek to Foster and Protect Real-Time Reliability	9-22
16. Unbundled Generation Can Increase Predictability, Transparency and Self- Organization in Real Time	9-23
Conclusion	9-24

10 RECOMMENDATIONS	10-1
The Challenge of Reliability.....	10-2
The Neglected Professionals	10-3
Building Analytic Capacity	10-5
The Way to Go	10-6
Three Examples of "Reliability Analysis".....	10-7
Promoting Careers for Reliability Professionals.....	10-10
Improving Recruitment	10-11
Creation of the California Electricity Reliability Network.....	10-12
Conclusion	10-13
A SHORT HISTORY AND INTRODUCTION TO THE CALIFORNIA ELECTRICITY RESTRUCTURING, CRISIS AND SYSTEM.....	A-1
California's Pre-Restructuring Electricity Sector	A-1
Restructuring the Industry (1994-1998).....	A-4
The New Structure	A-7
The First Years of Operation (1998-2000).....	A-11
California's Electricity Crisis (2000-2001)	A-13
Operations In and Around the Control Room of the California Independent System Operator (ISO)	A-22
The ISO Control Room Meets the Electricity Market	A-25
B PRIMARY FEATURES OF HIGH RELIABILITY ORGANIZATIONS.....	B-1
High Technical Competence	B-1
High Performance and Close Oversight	B-1
Constant Search for Improvement.....	B-2
Hazard-Driven Adaptation to Ensure Safety	B-2
Highly Complex Activities.....	B-2
High Pressures, Incentives and Shared Expectations for Reliability	B-2
Culture of Reliability	B-3
Reliability Is Not Fungible.....	B-3
Limitations on Trial-and-Error Learning (Operations Within Anticipatory Analysis)	B-3
Flexible Authority Patterns	B-4
Positive Design-Based Redundancy	B-4

C REFERENCES.....	C-1
--------------------------	------------

LIST OF FIGURES

Figure 2-1 California Electricity System	2-6
Figure 2-2 The HRN as the Organizational Network Connecting Markets and Technology	2-7
Figure 2-3 Performance Conditions for the Focal Organization, the ISO	2-11
Figure 2-4 Illustration of Operator Experiences of Performance Conditions	2-14
Figure 2-5 Recorded Transitions Between Performance Conditions	2-17
Figure 2-6 Breakdown of Estimated Percent Time Spent in Each Performance Mode, June 2001 - May 2002	2-20
Figure 3-1 Typology of “Just-In-Time” Behavior of Control Room and Wraparound Staff	3-11
Figure 4-1 Four Types of Always-On ISO Management Under Real-Time Performance Conditions	4-5
Figure 6-1 Fungibility of Reliable Electricity Over Time	6-7
Figure 8-1 Divergent Forces Redefining Reliability	8-2
Figure 10-1 Reliability Space and Key Professional Activities	10-4
Figure A-1 Average Electricity Rates 1965-1980 (in Nominal Dollars, cents per kWh)	A-2
Figure A-2 Average Electricity Rates 1982-1991 (in Nominal Dollars, cents per kWh)	A-3
Figure A-3 California’s New Electricity Market Structure Established by AB-1890 Source: Marnay (<i>et al</i>) (2001), p. 6, Figure 4	A-8
Figure A-4 California’s Market Design and Market Reality in 2000 Source: Lee (2001), pp. 4 and 6, Figures 5 and 9	A-16
Figure A-5 Average Prices (in Dollars) Utilities Paid for Electricity in the Cal PX Day- Ahead Auctions Source: Congressional Budget Office (2001), pp.18, Figure 4	A-17
Figure A-6 A Schematic Display of the ISO’s Folsom Control Room	A-24

LIST OF TABLES

Table 2-1 Principal Features of High Reliability Organizations (from 1980s Research)	2-3
Table 2-2 Major Explanations, Proposed Solutions and Principles for Reliability in Response to the California Electricity Restructuring and Crisis	2-4
Table 2-3 Performance Modes Summarized	2-16
Table 4-1 Percentage Time Breakdown of Shift Manager Tasks.....	4-4
Table 4-2 Feedback and Discovery of Error in the ISO	4-8
Table 5-1 Positive and Negative Implications of Persistently Incomplete Design	5-11
Table 6-1 Fungibility of Service and Grid Reliability	6-10
Table 8-1 Comparison of Selected HRO and HRN Features.....	8-14
Table 9-1 Principal Push and Pull Factors to Real-Time Reliability	9-3
Table A-1 A Chronological Outline of California’s Electricity Restructuring Process	A-10
Table A-2 Merchant Generation in California	A-13
Table A-3 Volume Purchased Out-of-Market June through December 2000 (in Megawatts)	A-17
Table A-4 Comparison of California’s Stage 1, 2 and 3 Emergencies	A-19
Table A-5 California’s Rolling Blackouts	A-20
Table A-6 California’s Load Growth and Operating Reserves (in Megawatts)	A-21
Table A-7 Conservation in California During the Summer of 2001 (in Megawatts)	A-21
Table A-8 An Overview of the Functions Within the ISO Control Room Teams.....	A-23

1

INTRODUCTION AND ROADMAP

Introduction

The restructuring of California's electricity system has been the subject of great comment and agitation in California and beyond. The creation of wholesale electricity markets has been analyzed by economists, regulators, policy-makers and pundits alike. The topic of bar talk and committee hearings, the crisis induced by restructuring has held the rapt attention of Californians as citizens and consumers.

Yet for all the attention and analysis, the most critical component of electricity restructuring and its subsequent crisis has been, we believe, ignored when not misunderstood. This is the critical challenge posed by restructuring to the very reliability of the electricity generation, transmission and distribution system that restructuring sought to improve. Our report brings a new reliability perspective to bear on restructuring and its continuing aftermath. The reliability we are focusing on has two dimensions: the safeguarding of the physical grid—the wires, switches and hardware through which electricity transmission occurs—and the constant availability of power to customers, the "always-on" service demanded by the consuming and voting public.

The report addresses and makes recommendations about unavoidable obstacles restructuring poses for both reliability dimensions. The problems were virtually unforeseen by the architects of electricity reform, and they are largely ignored in the debate about further restructuring and redesign. What our research reveals is a system design, which—in the view of the major theories of and approaches to service reliability—should hardly be reliable at all. Yet the restructured electricity network has achieved high levels of reliability in ways which confound theory and which enlarge our understanding of grid and service reliability.

Specifically, we examine the strategies and design implications of operators in and around electricity control rooms who are increasingly preoccupied with "real-time" planning, that is, for the hour ahead or within the current hour for operations. The implications of the reliance on real-time operations for the high reliability provision of electricity (i.e., for providing electricity safely and continuously even at peak load requirements), are, we have found, particularly not well understood by state and federal regulators and policy makers.

For many grid operators and managers, the all-absorbing preoccupation with real-time operations is what defined the electricity crisis. The report argues the proportions of time spent between real-time operations and longer-term management, planning and investment have permanently changed because of the California electricity restructuring and the crisis it induced. Adjusting to

this new real-time burden and balancing it with other strategies for grid management is now, we argue, an enduring feature of the level and nature of electricity reliability in California.

Our report examines the hitherto untold story of the California electricity crisis—namely, what it looked like from within the control rooms of those who kept the lights on for the California public and who, for that matter, prevented system failure of the western grid. It was in the control rooms of the California Independent System Operator (ISO) and those of the generators and distribution utilities where we saw and still see real-time operations in its most persistent form. Nowhere else do you observe the permanent changes being made in our notions of grid and service reliability and how operators must manage for them.

Two cautions are in order before proceeding in the report. While the consequences of California electricity crisis of 2000-2001 continue to find their way into today's headlines, analysts and researchers now have some distance to ask and answer the big questions:

1. Why did it happen?
2. Why was it so difficult to prevent blackouts, even when the state was willing to pay for reliable electricity at almost any price?
3. How did the control rooms actually keep the lights on virtually all the time during the crisis, albeit "just-in-time" and with not much time to spare?
4. How can we expect private-sector generators, public-sector regulators, and not-for-profit grid managers to provide electricity reliably in the future, when they have the competing goals and interests instituted through restructuring and manifested in the crisis?
5. What are the ongoing consequences of the electricity deregulation and crisis for California?

Our report touches on all five questions, but is primarily devoted to the third and fourth. The report is not an exhaustive investigation of California electricity deregulation, or more properly, electricity restructuring (a short summary of the California electricity restructuring and crisis is provided in Appendix A). Nor is this report about how to improve the long-term planning and investment in the grid, and the policy and institutional strategies that are needed to get there.

However, as the report is about the institutional design implications of the changing nature of electricity reliability under the crucible of events like the California electricity crisis, we have a great deal to say about the restructuring and transformation of the integrated utilities into an untested network of private generators, non-profit entities, and regulated distribution utilities.

Chapter 2 of the report lays out the context and framework for analyzing California's electricity restructuring and crisis from the perspective of grid and service reliability under persisting real-time performance conditions. Chapters 3 through 9, describe seven under-recognized system features and challenges core to the organizational requirements and performance conditions for "always-on" reliability. A chapter is devoted to each feature and our detailed case studies of the features draw on multiple sources (interviews, documents and observations) gathered during the recent crisis.

The features and challenges laid out in Chapters 3 through 9 are meant to inform and improve the advice of the key professions involved in the design, planning and operation of the electricity sector (economics, engineering and politics). These disciplines are and will continue to be the main source of institutional redesign proposals making their way up on the radar screen of policy and management attention. We believe Chapter 2 is the heart of the report because its findings have important repercussions for such proposals. Clearly, the professions themselves are the first to identify and address these implications on the basis of their expertise.

As a way to jump-start professional rethinking about electricity reform, Chapter 10 draws out what we see to be the major implications that flow directly from our findings in previous chapters regarding the organizational and institutional requirements to be addressed by the major disciplines working to improve the design and operation of the California electricity system. If our many interviews are the guide, real-time reliability requirements will persist into the future. They are not an anomaly and they have major management requirements for the future. Indeed, the 2000-2001 crisis is prologue to future volatility, the unpredictability and uncontrollability for which our policy makers and public managers can and must be better prepared.

We come at these topics from a disciplinary perspective different than economics, engineering and politics. Our perspective is that of organization studies, more specially, the current approaches to organizational reliability. Our framework and analysis, nonetheless, is intended to be of major import and interest to economists, engineers, policy-makers and their policy analysts alike. The report's perspective and views are unfamiliar and some may be controversial, but all represent a call for facing up to the high costs and consequences for having made electricity reliability such a precondition for our lives and livelihoods.

Before describing the research methods used in this report, we must address, at least in passing, the aforementioned questions that are not the primary focus of this report. Since we seek to identify the institutional and management requirements of ensuring high reliability of electricity provision within the network of organizations put in place by restructuring, it is necessary to say something about restructuring and deregulation generally, if simply to position the report's subsequent discussion.

The California Electricity Crisis

Fortune magazine placed the California electricity crisis and the related Enron collapse right after “9.11” and economic recession as the top major events of 2001, arguing that “the [California electricity] crisis left the deregulation movement in tatters.” Note the word “related” in the preceding. It is our view that deregulation and the Enron collapse—and with it the tsunami impacts on Enron *et al* (Dynergy, Williams, Calpine, Mirant, Duke, others)—are part and parcel of the same crisis of energy markets that act as if high reliability doesn't matter unless “the price is right.”

The California electricity crisis was not just a multi-billion dollar sink-hole—\$13 billion dollar for the utilities and \$9 billion dollar for the state, and still counting—with budgetary and economic ramifications that continue to be felt in the state and nationally. Depending where you stand, the widening California energy market crisis and the Enron *et al* collapse look to be a

failure in accounting, dot coms, or deregulation. Actually, it is all three. Enron *et al*'s initial "success" dramatically increased volatility of energy prices and supplies. Yet increased volatility without options to respond to it is antithetical to consumer demands for service reliability from their critical infrastructure of electricity. Energy trading thrives on price volatility and that volatility is the enemy of real-time reliability when balancing load and generation in the control rooms.

The parallels between the California electricity crisis and the crisis of Enron *et al* are remarkable. Early on, both the ISO, charged with ensuring the high reliability of energy transmission in the state, and Enron epitomized energy market deregulation. Both have had start-up cultures that value risk-taking and change, but only one survived—namely, the one that acted as if reliability matters. Huge bankruptcies and failures—Enron last year, Dynegy last week—are now a feature of the volatile energy sector, Enron of course being the second biggest in U.S. corporate history and that of Pacific Gas and Electric (PG&E) the biggest in the history of U.S. utilities. To this list can be added California's market scheduler of electricity, the Power Exchange (PX), and who knows what private energy suppliers tomorrow.

The lack of transparency is another parallel. Under deregulation Enron became so complex that even senior executives did not have a full picture of the company, and the same certainly can be said of the deregulated energy sector in California, with its dispersed network of private generators and energy traders, non-profit ISO, and investor-owned utilities. Nostalgia for the regulated, integrated utilities model also grew during the California electricity crisis—vide all the calls for "re-regulation"—just as we heard calls for Enron *et al* to restructure themselves around the older regulated business of pipelines and power plants.

Yet it is the ISO and its reliability network that have survived. Why?

First, according to economists, reliability is only one attribute of electricity and can be traded off against other attributes, like how clean the energy is. Not so. Reliable electricity has become a critical infrastructure, and we as a society are unwilling to trade off the reliability against electricity's other features. That reality, as we shall see, becomes more and more unavoidable the more grid operators work under real-time performance conditions, that is, the closer they get to pulling things together "just-in-time," hour by hour, up until the last minute. Markets, instead, trade off everything, whatever time it is. Consequently, given society's unwavering demand for always-on electricity, we can expect reliability operations to be increasingly real time, since that is where market pressures to trade-off reliability against price are their weakest and most resisted.

Second, it was argued that deregulated energy markets would significantly reduce bureaucracy. Not so. As we shall see, it merely displaced those costs. Deregulated markets have created a far greater task of coordinating interdependency among the many market participants. The number of employees at the California ISO far exceeds those originally planned. It turns out a great many programmers for the market software and lawyers for the contracts are required under deregulation. It will continue to be harder to manage markets than it was to manage regulation, we predict, for reasons outlined in this report.

Third, it is true that the retail market for electricity was not deregulated the way the wholesale market was in California. But to say that deregulation was not really tried is a bit like saying that African socialism or the Chinese Cultural Revolution did not work because neither was really given a chance. Very smart and dedicated people over a sustained period of time gave deregulated energy markets their best shot and there were still a variety of unforeseen and undesirable consequences. From the perspective of our report further deregulation raises very real prospects for additional problems, not least of which is increasing system volatility while at the same time pauperizing the options needed to deal effectively with the volatility in order to have grid and service reliability.

Fourth, the real experiment in energy markets has not been deregulation but something even more far-reaching. It has been the concerted assault on an extraordinary set of high reliability institutions, PG&E and Southern California Edison (SCE), on the premise that they can be wounded while still maintaining if not actually improving grid and service reliability in the process. Our report draws out the implication of this stupendous conceit.

What has worked in California but not for Enron *et al* is simply this. The generation assets sold off under electricity restructuring were worth so much precisely because they were backed up by a governmental willingness to ensure that reliability of electricity would not suffer as a consequence. The experiment of not taking reliability seriously has led Californians and others to pay a very high price to make good that guarantee. All of this could have been done more efficiently with forethought to institutional design and transaction costs.

It is said of Americans that they hate regulation but demand all manner of regulations to ensure their own personal safety, health and well being. Now we can see that our critical infrastructures are a matter of the latter, not of the former. Americans, not just Californians, have shown themselves willing and able to spend billions upon billions in the name of ensuring the reliability of certain critical infrastructures—think only of Y2K. Such transfers of income demonstrate that critical service reliability is not like any other "attribute" we know. In fact, it is more of a state condition than an attribute, in the same way that a society's distribution of assets is treated as a given from which any set of prices are established. Substantially change asset distribution and you change the prices; substantially change grid and service reliability and you change the assets.

All this can be summed up as the Law of Reliability Matters: The more services demanded from a single resource, such as electricity (or water or telecommunications or transportation...), the greater the demand for the high reliability management of those services and that resource. Critical infrastructures are deliberately designed so that we, as a society, depend on them to such an extent that it is worse not to manage them than to manage them. They require highly competent, always-on management, the net benefit of which exceeds what would be the case if we were not to manage these structures as reliably. Electricity "restructuring" is thus an appropriately named challenge, for it affects society as a whole. The following chapters and recommendations speak only to a part of that challenge, but it is a challenge deserving much more attention by all those who take not only just real time but also the long-term seriously.

Study Project, Research Question and Methods

The report is based on research undertaken as part of the Networked Reliability Project, a joint initiative involving faculty and researchers from Mills College in Oakland and Delft University of Technology in the Netherlands. The principal researchers are: Michel van Eeten, Project Leader, Associate Professor-Delft; Emery Roe, Professor of Public Policy-Mills; Paul Schulman, Professor of Government-Mills; and Mark de Bruijne, graduate student-Delft. Schulman was involved in the study of high reliability at PG&E (including Diablo Canyon) in the late 1980s (e.g., Schulman, 1993a, 1993b, 1996), while Van Eeten and Roe have studied high reliability issues related to water and power at the California Department of Water Resources (CDWR) and the Bonneville Power Administration (BPA) in the late 1990s (van Eeten and Roe, 2002).

The Project's key research question is: How can critical infrastructures, such as electricity, water supply and telecommunications, ensure and maintain high service reliability under rapidly changing conditions? Developments, including deregulation, conflicting reliability mandates, technological innovation and rapidly growing demand, have had one common denominator: They make highly reliable service provision of peak load services safely and continuously through time more and more dependent on a widening set of organizations operating within the infrastructures and the network industries associated with them.

The Project examines the challenge and responses through a series of case studies on different critical infrastructures whose common theme is the limited potential for top-down coordination to achieve reliability. We are studying three critical infrastructures within the context of deregulated institutional and market structures: electrical power, telecommunications, and water supply systems. Other areas to be examined include financial services.

As of the date of writing, 60 interviews for the California case study have been complete:

- Thirty-three in and around the ISO control room
- Eight at PG&E (e.g., in and around their Transmission Operations Center and the Operations Engineering units)
- Five with a large generation supplier (a senior generation official and control room operators in one of its California large plants) and a private market energy trading dot.com
- Fourteen others (including in the Governor's Office, California Public Utilities Commission (CPUC), California Energy Commission (CEC), Electric Power Research Institute (EPRI), Lawrence Berkeley National Laboratory (LBNL), University of California, Berkeley (UCB), and Stanford

Most interviews were for an hour or longer and undertaken between April – December 2001. We thank the interviewees here for their cooperation, candor and time, often during the height of the California electricity crisis. As is the convention in social science and public policy research, confidentiality and anonymity are maintained in the text of the report, such that all quotes from the interviews remain unidentified with respect to interviewee and his or her title. In all cases, those quoted have been given the opportunity to amend their statements, and we have honored their requests whenever made.

Roadmap

Chapter 2 sets out the overall empirical context and conceptual framework for the report. We discuss the major theoretical and professional explanations for why service reliability was difficult to achieve under restructured and crisis conditions. The real surprise, however, is that service and grid reliability was much better than the major theories and professions would lead us to believe. We set out a conceptual framework based on our interview data to explain how this could be. In particular, we show how a high reliability network (HRN) of generators and utilities, with the California Independent System Operator (CAISO or ISO) as its focal organization, was able to adapt to a variety of different performance conditions in order to balance load and generation under variable conditions of system volatility and network options. Under conditions of high system volatility with variable options and strategies (ranging from high to low) to respond to that volatility, we observed "just-in-time" and "just-for-now" performance modes, or what we call "real-time reliability." The other two performance modes we observed found under conditions of low system volatility and variable options were "just-in-case" and "just-this-way" performance.

Just how do “system volatility” and “network options variety” work in practice? What does "just-in-time" and "just-for-now" performance look like from inside the control rooms of the HRN? For that matter, what do control room operators actually have to do to keep the lights on and the grid safe? In answer to these and like questions, Chapters 3-9 examine the important features and reliability challenges to the California electricity system through our interviews, observations and research. The design features are important both for maintaining real-time reliability (that is, "just-in-time" and "just-for-now" performance under conditions of high system volatility such as induced by restructuring) and for posing challenges to the improvement of that reliability. They are connected to the Chapter 2 framework and flesh out the reliability challenges faced on the ground in the HRN. Each of the following chapters discusses a feature that we believe has implications that are often misunderstood, when not overlooked.

It would be tempting to fashion this report as One Big Story in seven parts: Challenges (posed by incomplete design, always-on service and non-fungibility of reliability) lead to Responses (redefining reliability, markets and fallbacks, and wraparounds) with one major Consequence (push and pull to real time). That, however, would be too artificial. First, a number of cross-cutting issues, particularly the importance of emerging informal networks, connect the chapters and we return to them continually throughout. Second, the features and challenges describe not so much one story as different components of the same California electricity system. Thus, it would be better to see chapters 3-9 as an interrelated set of themes, each of which requires attention and has implications for institutional design and reliability improvements.

Chapter 3 is about a frequently misunderstood feature and challenge—the wraparound, particularly that for the ISO control room. The importance of a control room’s wraparound derives from the fact that it addresses both system volatility and options variety at the same time. Control room support staff seek to buffer the control room from system volatility, while at the same time assist control room operators in terms of developing and implementing various options to respond to that volatility so as to balance load and generation. Chapter 3, “The wraparound of market and technology,” reveals the substantial organizational infrastructure of

programmers, engineers, lawyers, economists, and others needed to enable the electricity market to function and connect its outcomes to the physical properties of the grid, including buffering the technical system against market volatility. In the case of the ISO, the support infrastructure is the “wraparound” of rooms and staff adjacent to the control room, in which the three main market desks exist (day ahead, hour-ahead, and the real-time imbalance). It takes a great many people to enable markets to function as an “invisible hand,” and the existence of extensive wraparounds was a significant factor contributing to real-time reliability of electricity before, during, and after the electricity crisis. Change these wraparounds and you change the performance conditions under which the HRN operates and the adaptability of the focal organization, the ISO, to respond to these changing conditions.

Real-time reliability also depends on the interaction among all the control rooms and their respective wraparounds across the HRN, that is, between the ISO, private generators and their security coordinators (SCs), and the distribution utilities. This HRN-wide interaction derives from the always-on feature of electricity.

Chapter 4, “Electricity as an always-on service,” highlights the reliability consequences of this feature. The “always-on” requirement means that feedback on performance—on whether load and generation are balanced—must be immediate, continuous and relatively unambiguous to parties within the HRN. “Immediate” means the feedback is real-time information, “continuous” means the feedback is always there 24/7, and “unambiguous” indicates that the feedback is focused specifically on or related to balancing load and generation. When information is continuously immediate and unambiguous in terms of reliability performance, formal and informal processes of alertness, readiness and accountability develop with respect to both one’s own actions within the HRN and the trustworthiness of others in the network. Parties in the HRN build up knowledge of who can be trusted to be ready to do what, where and when. Even though a “culture of reliability” is not dominant in the network-wide provision of electricity (as it did in key parts of the earlier integrated utilities), we did find official and unofficial systems of shared knowledge, timeliness and accountability for actions and their consequences. These help compensate for information asymmetries and design incompleteness in the HRN that promote strategic behavior and gaming. In other words, being alert, ready and accountable increased the options variety and helped ensure the network was as reliable as it was during the crisis.

Wraparounds and systems built on immediate, continuous and unambiguous feedback can only do so much, since they do not control system volatility nor determine the full range of options available to respond to that volatility. A great deal of the latter is outside the operators’ hands, as it has always been. Chapter 5, “Persistently incomplete design,” questions the prevailing view that the electricity crisis was the result of poorly designed institutions and that once the missing elements are provided (e.g., more generation or retail market deregulation), the California electricity system will become sufficiently stable and predictable. For those who hold such a view, the problem today looks to be incomplete regulatory oversight, gaps in long-term planning, and misaligned business models. Our research, however, reveals two major mechanisms that underlie and maintain an enduring incompleteness of policy and institutional design in the restructured California electricity system: the overall complexity and the internal dynamics of any such designed network. The conclusion of the chapter is that design incompleteness is permanent and that improved mechanisms to cope with it are essential to ensuring reliability

under enduring real-time performance constraints. Chapters 6 through 9 describe four features which are increasingly important for HRN in “pulling things together hour by hour up until the last minute” in order to ensure real-time reliability of electricity. First, in order to have electricity be always on it is treated by the ISO—more specifically its control room operators—as non-fungible in real time. Chapter 6, “Non-fungibility of reliability in real time,” describes how the demand for reliability has rendered electricity into not “just another commodity”—which has been a basic assumption behind deregulation.

Contrary to the economics of deregulation, our interviewees demonstrate that reliability of electricity is not a quality-of-service attribute that can be substituted by other attributes at the margin, e.g., by trading off reliable electricity for “green” electricity. There is a point at which dollars no longer can replace reliability, i.e., it is cheaper to provide the service reliably than it is to pay people to forgo that service. Indeed, the only time service reliability becomes fungible is when it directly jeopardizes the grid. In Chapter 6 we show that grid reliability is non-fungible to such an extent that the state was willing to trade off service reliability (keeping the lights on) in those few instances where shedding load and controlled blackouts were needed to preserve that reliability in the immediate term. The non-fungibility of grid reliability in real time is a primary reason why California electricity was provided as reliably as it was during the crisis, even under market meltdown and controlled blackouts.

Chapter 7, “Markets and their fallbacks,” is heavily based on our interviews and observations and indicates how the ISO control room and its wraparound actually cared for markets and market transactions even at the height of the crisis and also undertook other interventions that compensated for market problems and failures. When the day-ahead, hour-ahead and real-time market desks were breaking down, the ISO managed to fill the coordination gap left behind in order to maintain grid and service reliability. All kinds of market transactions between HRN partners—including those paradoxically named “out of market” transactions—kept the electricity reliability as high as it was. We also found non-market fallbacks for the failing markets, such as informal network coordination and organizational adaptations. This ability to “carry the market on by other means” contributed significantly to the relatively high levels of real-time reliability observed during the crisis.

Chapter 8, “(Re)defining reliability,” addresses the fact that definitions of the reliability of the system are changing. Our interviewees tell us that there is not one official reliability standard that has not been pushed to its limits and beyond in the California electricity restructuring and crisis. Our findings run counter to the belief that such a breaching of standards looks to be *prime facie* a decline in reliability. An important part of this redefinition, we argue, is adapting the standards to conditions of high volatility—in the words of one interviewee, to give the ISO a “bar that they can jump.”

The above system features alone do not describe how the ISO and other parties in the HRN have come to find themselves increasingly maintaining reliability under persisting system volatility and changing performance conditions. Chapter 9, “The push and pull to real time,” summarizes sixteen factors at work in pushing and pulling the balancing of load and generation together at the last minute. Partly, these factors illustrate the aphorism that necessity fosters invention. While the original design for the HRN envisioned that real-time operations would deal with the

last few percent of electricity transactions, a large portion of overall electricity demand was provided hour by hour or at the last moment during the crisis (Appendix A). This has been considered a failure in reliability by many insiders, who feel pushed into real time as the last resort to prevent grid collapse. Our report underscores not just the push but the pull factors as well driving the California electricity system into real-time, last-minute provision of services; factors that will persist notwithstanding new generation and new market structures to come on line in the future.

Finally, Chapter 10 sets out the recommendations of our report. They may also set this report apart from those that have preceded and will follow it. We have no administrative reform, policy initiative, sector restructuring, technological fix or institutional redesign to commend. Our recommendation is, we believe, being considerably more useful for improving the reliability of grid and service in California: Enhancing the analytic capacity, career development, and recruitment of the reliability professionals at the heart of our electricity critical infrastructure. In particular, we argue that all proposals to improve the California electricity system should pass a reliability-matters test and we give three examples of what we call “reliability analyses,” including analyses of real-time residential metering and homeland security recommendations made with respect to the electric grid.

2

CONTEXT AND FRAMEWORK

Context

The most important feature of the California electricity crisis is that no one really predicted it. Yes, some saw elements of it, and with that exact science of hindsight there are those who now say they saw it coming from the beginning. Not one of those we interviewed predicted the crisis that ultimately did unfold, however. By the end of the report it will be clear why.

Here, though, another failed prediction matters. Nothing we know about organizations would have forecasted that electricity was provided as reliably as it actually was during the crisis. In fact, current organization theory predicts that restructuring of the electricity sector should have reduced reliability considerably more than was actually the case.

Clearly, the threat of blackouts was very real. The number of declared Stage 3 emergencies mushroomed from none in 1999 and one in 2000 to 38 in 2001 (trend lines were similar for Stage 1 and 2 declarations). Balancing load and generation was also a heroic effort achieved at times only in the last minute. Yet, all that said, actual blackouts were few.

Notwithstanding the popular view of rolling blackouts sweeping California during its electricity crisis, in aggregate terms—both in hours and megawatts (MW)—blackouts were minimal. Rolling blackouts occurred on six days (Jan 17-18, March 19-20, and May 7-8 inclusive) during 2001, accounting for no more than 30 hours. Load shedding ranged from 300 to 1000 MW in a state whose total daily load averaged in the upper 20,000 to lower 30,000MW. The aggregate amount of megawatts that was actually shed during these rolling blackouts amounted to slightly more than 14,000 Megawatt-hour (MWh), the rough equivalent of 10.5 million homes being out of power for one hour in a state having some 11.5 million households—with business and other non-residential customers remaining unaffected. In short, the California electricity crisis had the effect of less than an hour's worth of outage for every household in the state. These and like figures are discussed at length in Appendix A.

The level of high reliability observed is unexpected in light of the dominant theories of reliability in large technical systems. Upon closer inspection we have identified a reliability quite different than anticipated in that theoretical literature. This report is about that specific networked reliability in the California electricity system and why it matters.

A dominant theory of reliability is termed Normal Accidents Theory (NAT). Developed by sociologist, Charles Perrow, and his colleagues, NAT argues that tightly-coupled technological systems, such as electricity grids, are inherently prone to accidents that can bring the whole grid

down (Perrow 1999 [1984]). Consistent with that perspective, the history of electricity and drive for greater reliability are pock-marked with large-scale, widespread blackouts—East Coast, West Coast, and In-Between—whose collapses and near misses compelled institutional change to improve the reliability of the grid (e.g., Hauer and Dagle, 1999).

Yet if the potential for cascading accidents is an inherent feature of large complex systems, such as the tightly-coupled, regional power grids across the states, why haven't there been more accidents? More formally, how do some systems, with complex technologies and in unstable task environments, still manage to meet peak load service production in a continuous, safe fashion? High Reliability Theory (HRT) arose in part as an answer. HRT was developed in the late 1980s by organization theorists interested in how complex organizations and institutions maintain their activities in situations where failure, error, and accidents are highly consequential (Rochlin 1996; La Porte 1996; Schulman 1993). High reliability organizations (HROs) studied included nuclear power plants, air traffic control, large water supply systems, hospital intensive care units, fire fighting units, and naval air carriers, among others. While HRT theorists can never fully explain high reliability to the satisfaction of NAT proponents, the former find themselves in the role of Galileo to the latter's Cardinal Bellarmine: *E pur si muove* (And yet it moves).

Both NAT and HRT would predict the restructuring of the California electricity sector should have noticeably undermined the reliable provision of electricity. For its part, NAT would see the coupling of the PG&E and SCE grids into a newly configured grid, with altogether novel and higher flows of energy as an increase in the technology's tight coupling and complex interactivity. The probability of cascading failures should have increased accordingly.

For its part, HRT would have come to a similar conclusion, but by a different route. HROs, including some of the older integrated power utilities, are almost exclusively preoccupied with ensuring stability of internal processes, which, if not managed, greatly magnify consequences for error. Consequently, they seek to stabilize both inputs and outputs, much as manufacturing firms might seek a vertical integration of upstream and downstream elements in production and marketing in order to stabilize the manufacturing process. Among HROs, eleven interrelated features were found crucial to maintaining reliability in the face of highly consequential hazards, such as those found at PG&E's Diablo Canyon nuclear plant. These features are summarized in Table 2-1 and detailed in Appendix B.

The features are not sufficient conditions for high reliability. Nor is there a formula and there are certainly no guarantees that, even if the features were all present, one would find highly reliable service provision as a result.

It is best to think of the eleven features as probable, necessary conditions: Where there was high reliable service provision—at least during the late 1980s— these features were also generally present. Recent research (Van Eeten and Roe, 2002) found the above features also in and around the control rooms of HROs of large water supply and hydropower systems. These rooms are the one place where we observed the technical competence, complex activities, high performance at peak levels, search for improvements, teamwork, pressures for safety, multiple (redundant) sources of information and cross-checks, and the best example of the culture of reliability, all

working through technologies and systems that build in sophisticated redundancies to buffer against potential failure and catastrophe.

Table 2-1
Principal Features of High Reliability Organizations (from 1980s Research)

- high technical competence
- high performance & close oversight
- constant search for improvement
- often hazard-driven adaptation to ensure safety
- often highly complex activities
- high pressures, incentives & shared expectations for reliability
- culture of reliability
- reliability is not fungible
- limitations on trial & error learning (operations within anticipatory analysis)
- flexible authority patterns
- positive, design-based redundancy to ensure stability of inputs & outputs

From this perspective, HRT would thus predict that restructuring of the electricity sector necessarily introduces substantial instability in electricity reliability by increasing the unpredictability of inputs and outputs and undermining the principal features associated with high reliability. Deregulation and restructuring, after all, unbundled generation, transmission and distribution of electricity, where what was once done by the integrated electricity utility is now provided by a network of organizations with competing, if not conflicting goals: profit for the independent generator, reliability for the not-for-profit ISO, and a life on the cusp of bankruptcy for the regulated utilities. Restructuring has meant that the control room operators are now dispersed through organizationally distinct units, with the plant generator linked to the company trading floor, the transmission center of the utilities dedicated to distribution only, and the ISO as the manager but not owner of the grid for ensuring service and grid reliability. What was once the culture of reliability is now a network of divergent interests bordering on open polarization, warns HRT.

While the California restructuring and electricity crisis brought together many forces conspiring against high service reliability, our research identifies emerging forces, strategies and arrangements that were effective in providing more reliability under extremely adverse circumstances than one would have predicted on the basis of what NAT and HRT alone. These emergent properties and challenges point to requirements for *networked reliability* in the electricity sector. The more important features and challenges are discussed in chapters 3-9. They, moreover, have implications for better economics, engineering, policy and bureaucracy in

meeting service reliability through critical infrastructures. These implications in turn are discussed in Chapter 10.

What must be highlighted in the remaining paragraphs is the import of the following chapters. The good news is that high reliability is possible in inter-organizational networks of divergent interests. The bad news is that this reliability is significantly different from what the economists, policy-makers and engineers have been telling us it is. Our findings differ considerably from the explanations of most experts, who assume that electricity services can be provided as reliably as you want—for a high enough price or given sufficient enough political will or with the best engineering possible.

The following table summarizes the major explanations and proposals concerning the California electricity crisis among those economists, policy-makers and engineers we reviewed and interviewed (Table 2-2). Appendix A provides more information on each.

**Table 2-2
Major Explanations, Proposed Solutions and Principles for Reliability in Response to the California Electricity Restructuring and Crisis**

Disciplinary Perspective of Interviewees	Major Explanation of the Crisis is...	Major Principle of Reliability should be...	Major Solution is...
Economics	Officials did not deregulate enough	It is what you will pay for	Officials must deregulate further, i.e., deregulate the retail market
Politics & Policy	Vested interests struck a compromise that resulted in poor institutional design for reliability through markets	It is always there as a right	Officials must redesign the governance and regulatory structure for electricity provision, e.g., more assurances and better assignment of responsibilities
Engineering & Technology	The grid as a technical system was not designed for the level and flows of electricity under market conditions	It is provided as needed	Officials must redesign the control structure and technology for the grid, giving priority to investing in and updating the grid according to best possible engineering practices

There is no presumption that these views characterize all of economics, policy making and engineering nor that the views of those interviewed for this report weren't more complex when we probed them over a longer period of time.

That said, the economists we canvassed by and large believe that reliability is what one would be willing to pay for under a fully deregulated system. The problem we now face in their view is to complete that deregulation. Engineers argue that reliability is best assumed through automated technology and systems drawing upon expert knowledge as and when needed. The problem in their view is that the grid and its control structures need to be improved to meet the new, deregulated conditions under which it now has to perform. Politicians and policy-makers often act as if reliability is a right that should always be there, while privately admitting that any real redesign will inevitably be a matter of unhappy compromise. If this political perspective is taken seriously, then the assurances about electricity being a right and responsibilities for ensuring its reliability must be assigned or in place before the eventual compromises proceed ahead.

Each narrative has merits, and the bulk of publications on the California electricity restructuring and crisis will doubtless be permutations of them. In what is fast becoming conventional wisdom, the interplay of the three narratives has their proponents and opponents agreeing only on one thing: The California electricity restructuring and crisis ended up a disaster story of reliable-electricity-at-any-price.

Our research, however, found an altogether different perspective with very different challenges for grid and service reliability. That perspective is summarized in the following framework, the important design features of which are teased and parsed out in subsequent chapters. The framework, which is derived from our interviews, is expressed in formal terms and guide the rest of the report.

Framework

From a reliability perspective, we describe the California electricity system in terms of three components (Figure 2-1): the reliability task environment (*RTE*) and within it, the regulatory reliability network (*RRN*) and the high reliability network (*HRN*). Together, the *RTE*, *RRN* and *HRN* constitute what we term the *California electricity system*.

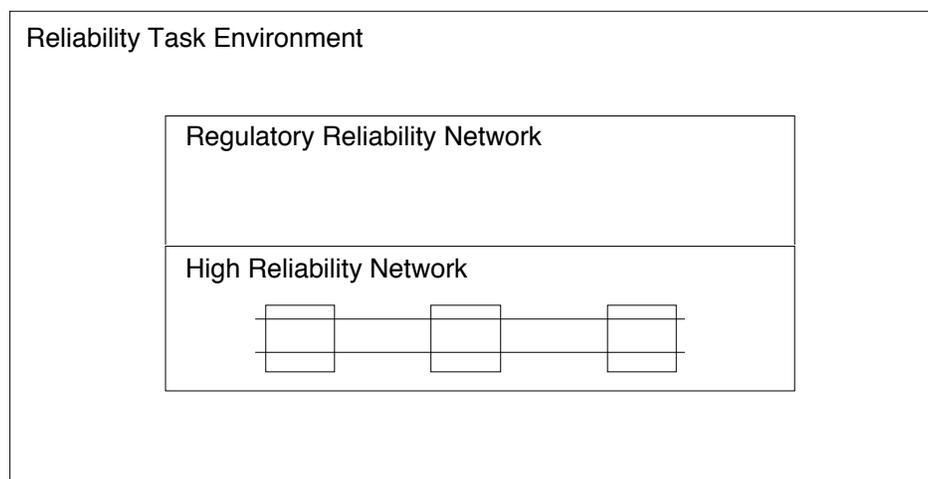


Figure 2-1
California Electricity System

What are the functions, duties and responsibilities of each system component? The details are provided in Appendix A, should the reader be unfamiliar with the terms, acronyms and organizations that follow.

Simply put, the HRN organizes for electrical service provision and delivery and includes the organizations and units that have direct operational responsibilities for this provision and delivery. In the California electricity system, the HRN consists of the control rooms and support staff of the ISO, the distribution utilities, the private generators along with scheduling coordinators (market traders and California Energy Resources Scheduling [CERS] division) and the adjacent control areas. The HRN is not a “high reliability organization” that stabilizes its inputs and outputs in order to ensure service provision, safely and continuously. Since the units in the HRN have divergent and, in some instances, directly competing interests, stability of inputs and outputs to ensure reliable electricity is a greater challenge for the HRN than it was for the earlier HROs.

The RRN sets the mandates and criteria for high reliability, i.e., it establishes and/or enforces standards that define reliability for units in the HRN. The standards include formal regulations under a government mandate and regulations developed and enforced by system participants themselves, such as WSCC (Western Systems Coordination Council) standards. The RRN does not organize the provision of services directly. In the California electricity system, the RRN consists of the CPUC, FERC (Federal Energy Regulatory Commission), WSCC, NERC (North American Electricity Reliability Council), CEC and EOB (Electricity Oversight Board), among others. While these agencies and units have reliability of electrical service as a mandate, they also have other mandates to reconcile under their respective budget constraints. For example, the CPUC also sets policy for telecommunications and sees itself as more “operational” than some of the other state regulators; the CEC is not just concerned about electricity reliability; and the WSCC has other grids to worry about than California’s.

The RTE sets the context for the reliability-related tasks of the RRN and HRN. The reliability task environment includes customers, voters, businesses, elected officials and the public. The customer has demands, expectations, and even certain contractual rights to receive service in a reliable fashion. A widespread notion in California is that electricity is not like any other commodity and that access to cheap electricity is an obligation, if not right. Clearly, the RTE, RRN and HRN are connected, as when the public’s pervasive notion of electricity as a right influenced much of the intervention of the Governor’s Office to have CERS fund the ISO’s provision of electricity during the California electricity crisis.

Since the report’s focus is on the institutional implications of real-time operations and provision of electricity on the reliability of that electricity, we turn now to a detailing of the HRN.

The High Reliability Network

The relationships between and among the elements of the HRN are schematically drawn in Figure 2-2. To start at the beginning, three nodes of activity in the HRN were unbundled as a result of electricity restructuring—from right to left in Figure 2-2: generation, transmission and distribution. For this figure, the generation node is exemplified by one of the “Big Energy Suppliers,” in this case its trading floor and one of its California-based generation plants. Were Figure 2-2 drawn realistically, it would show all other scheduling coordinators and the generation they represent (for a more detailed figure, see Appendix A).

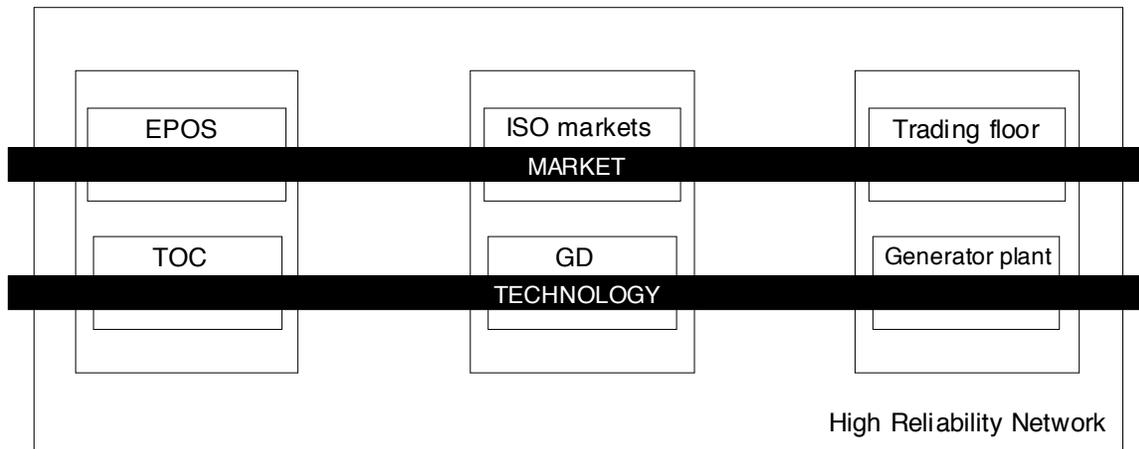


Figure 2-2
The HRN as the Organizational Network Connecting Markets and Technology

The transmission node is run by the market and grid transmission staff located in the ISO: respectively, its three market desks—day ahead, hour-ahead, and (BEEP) real-time imbalance markets—and the generation dispatcher along with grid support staff in the control room. The day and hour ahead desks are commonly associated with the congestion management and ancillary services market, while the BEEPer and gen dispatcher are commonly involved in the real-time imbalance market.

The distribution node in Figure 2-2 is represented by one utility—PG&E—in terms of its market and grid distribution staff, particularly its trading unit Electrical Portfolio and Operations Support (EPOS) and its distribution control room known as the Transmission Operations Center (TOC). A less schematic figure would also show the other distribution units, such as SCE, the municipalities ("munis") and irrigation districts, for example.

Restructuring has linked the three nodes in two ways: through markets and through the grid and its support technology. In theory, markets are to be the main coordinative mechanism for grid operations—that is, markets transactions should result in the electricity schedules that are to be the basis of grid operations. In terms of Figure 2-2, PG&E’s TOC, the ISO’s generation dispatcher, and the generation plant are organized around the grid, while PG&E’s EPOS, the ISO’s three market desks, and generation supplier’s trading floor are organized around market

operations. Reflecting these functional interconnections, market people tend to talk to market people and grid people tend to talk to grid people across the HRN's three unbundled nodes.

Many relationships are possible between and among the three nodes and their market and grid subnodes. In practice, formal relations are circumscribed by mandate, legislation and regulation. Within each node, market and grid operations are to be separate: EPOS is not meant to be in constant communication with PG&E's TOC; the generation plant staff are meant to have final authority over unscheduled outages not the supplier's trading floor; and the ISO's ability to undertake outage coordination activities with the generators is highly restricted by what types of information either node can obtain from the other. Informal communications outside formally restricted channels, consequently, continue to be important for ensuring the reliability of electricity in unpredictable situations, where salient information and communication are at a premium.

Because volatile (unpredictable or uncontrollable) situations increase the pressure to communicate between units, especially units within the same organization, an extensive set of support staff has grown up around the market and grid operations rooms of each unbundled node and its market and technology subnodes. We call this supporting infrastructure, the wraparound—in the ISO's case, the staff supporting market and grid operations are literally circled around the ISO's control room. Wraparound infrastructure supporting control room operations exist in other critical infrastructures, such as water (Van Eeten and Roe 2002). A wraparound is itself a site of formal and informal communication with counterparts in the other wraparounds, especially during crisis situations.

The respective control room and the wraparound form an organizational infrastructure around the market and the technology operations, the latter which we term the market matrix and technology matrix. Each matrix connects the market or the technology subnodes across the three nodes of generation, transmission and distribution. In Figure 2-2, the dark 'market' bar represents the flow of market transactions; the dark 'technology' bar represents the flow of electricity on the grid.

To a degree, the market and grid flows were also separated in the old utilities, The important difference between the HRN and the older, integrated, more HRO-like utilities, however, is that the markets are now supposed to coordinate across the matrices, with the focal organization for coordination being the ISO within a network setting. In reality, extensive organizational interactions between and among control rooms and wraparounds are needed for markets to operate, connect market transactions to the physical properties of grid and, ultimately, avoid grid collapse and islanding. The technology and market matrices are "opened" at their ends, as the California grid is itself part of the western grid and the "California" electricity markets themselves extend well beyond California state borders, e.g., several of the "big energy suppliers" are international in operations.

Three features of Figure 2-2 are to be especially noted. First, market transactions have to be connected to the physical grid, which requires constant non-market attention and intervention. The outcomes of the markets can easily cascade and undermine the California electricity system when generation does not follow load, congestion leads to overloads, or disruptions ramify into

high consequential hazards. Thus, the HRN is not only responsible for service reliability, but also for grid reliability. It is important to recognize that “high reliability in electricity provision” has two parts: reliability in the electricity as a service provided and reliability in the physical grid as the provider of that electricity. In addition to providing an always-on service, there is that other high reliability mandate of the HRN: to safeguard the physical grid itself, an enormously expensive physical asset which could be significantly damaged by HRN actions, including actions taken in pursuit of service reliability.

In this way, Figure 2-2 illustrates an important concept. The electricity markets only exist by virtue of a network responsible for grid and service reliability. The interdependencies of Figure 2-2’s subnodes, nodes, their wraparounds and the matrices are a necessary condition for the high reliability provision of marketed electric power. Reliable markets for reliable electricity do not, in other words, exist outside the context of the HRN. Based on our report’s findings, it is impossible to imagine any significant markets in electricity without significant HRNs to locate, house and support them.

Third, it is also easy to see how the HRN has the potential for cascading failure throughout the California electricity system: Actions in the HRN (that is, its nodes, subnodes, wraparounds and matrices) can immediately influence surrounding elements in complex ways, which spread through the system if not managed or averted. For example, the market strategy of a distribution utility might bring its own TOC into trouble or at least push the TOC to its limits. That market strategy (e.g., underscheduling, waiting until the last minute, or gaming the congestion market) can also disrupt ISO markets. These disruptions, in turn, can complicate matters for the ISO’s generation dispatcher in terms of real-time grid operations. In response, other generators may withhold bids in real time for strategic reasons, thereby having an impact back down to the level of actual plant operations. Such considerations constitute another reason why the network is “a high reliability network,” even though some of its participants have goals other than either service or grid reliability: Although its nodes are unbundled under restructuring, the prospect of cascades concentrates the mind wondrously, as Dr. Johnson once said of hanging. The issue, of course, is that not all the minds in the HRN are so uniformly focused in terms of reliability.

How Does the HRN Produce, Transmit and Deliver Reliable Electricity within the California Electricity System Under Varying Conditions?

This question is answered in detail in subsequent chapters, but for schematic purposes, the answer is arrived at by describing and explaining how the HRN balances load and generation continually under variable circumstances so as to provide reliable electricity to end-users. From our framework, balancing load and generation is the *reliability requirement of the HRN*. Each subnode and node, including their respective market and technology matrices, have, of course, many different tasks more or less requiring that overarching end.

More specifically, the following framework focuses on the match between, on one hand, the HRN options and strategies to balance load and generation and, on the other hand, the unpredictable or uncontrollable threats that arise to fulfilling the HRN reliability requirement of keeping that balance through time. A match results from having at least one option sufficient to maintain the balance under given conditions. That match means there is at least one available

network option sufficient to manage system variables in relation to their volatility. For example, price has been very volatile as a variable, thus requiring strategies and options for its management, such as caps, CERS purchases, or load reductions. In the absence of those options, volatility would threaten reliability even more. At any point, there is the possibility of a mismatch between the system variables that must be managed to achieve balance between load and generation and the options and strategies available for managing those variables. It should go without saying that the match is not automatic and requires management—in this case high reliability management within the HRN.

Within the California HRN, the mandate to balance load and generation, the core reliability task of the HRN, is located in the focal organization, the ISO. It is the only organization which simultaneously has the two reliability mandates: keeping the flow always on and protecting the grid reliably. This is what “balance” really means when the term is unpacked. Meeting the dual mandate involves managing the options and strategies which coordinate actions of the other nodes, subnodes and matrices of the HRN. As the focal organization, the options the ISO deploys are pre-eminently HRN-based or HRN-wide options, e.g., outage coordination is the responsibility of the ISO, but involves the other elements of the HRN.

Consequently, the ISO management can be categorized in terms of the variety of HRN-based options it, the ISO, has available (high or low) and the volatility of the California electricity system (high or low), resulting in four performance conditions and modes as set out in Figure 2-3.

		System Volatility	
		<i>High</i>	<i>Low</i>
Network Option Variety	<i>High</i>	Just-in-time performance	Just-in-case performance
	<i>Low</i>	Just-for-now performance	Just-this-way performance

Figure 2-3
Performance Conditions for the Focal Organization, the ISO

Volatility is the degree to which the focal organization, the ISO, faces uncontrollable changes or unpredictable conditions that threaten the grid and service reliability of electricity supply, i.e., that threaten the task of balancing load and generation. Many days have been those of low volatility, typically called “normal days” in the past. A clear example of high volatility are those days where a large part of the forecasted load had not been scheduled through the day-ahead desk, which means for the ISO actual flows are unpredictable and congestion will have to be dealt with hour by hour or at the last minute.

Options variety is the amount of HRN resources, including strategies, available to the ISO to respond to events in the system in order to keep load and generation balanced at any point in time. It can be approximated with conventional engineering parameters, including available operating reserves and other generation capacity and available transmission capacity. High option variety means, for instance, that the grid has more than regulatory required resources available (there are high or wide margins), low options means the resources are below requirements and, ultimately, that very few resources are left (low or tight margins).

The two dimensions of system volatility and network options together set the conditions under which the ISO has to operate and demand different performance models for reliability: “just-in-case,” “just-in-time,” “just-for-now,” and “just-this-way.” “Low” and “high” are obviously imprecise terms, though they are the terms used and commonly recognized by many of our ISO interviewees. In practice, the system volatility and options variety dimensions should be better thought of as gradients without rigid high/low cut-off points. Let us turn now to a brief description of each performance mode.

“Just-in-Case” Performance, Redundancy and Maximum Equifinality

When options are high and volatility low, “just-in-case” performance is dominant because of high redundancy. Reserves available to the ISO are large, excess plant capacity exists at the generator level, and the distribution lines are working with ample backups, all much as forecasted with little or no volatility (again, unpredictability and/or uncontrollability). More formally, redundancy is a state where the number of different but effective options to balance load and generation is high relative to the market and technology requirements for balancing load and generation. There are, in brief, a number of different options and strategies to achieve the same balance. The state of high redundancy is best summed up as one of maximum equifinality,

i.e., there are a many means to meet the reliability requirement, “just in case” they are needed. (Note that “equifinality” does not mean that the means are equal in all respects, only that they are roughly equivalent in realizing the end sought.)

High redundancy does not mean maximum reserves. Prior to restructuring, operating reserves were at a much higher percentage than they currently are (Appendix A), such that “just-in-case” performance is not the same state of performance that existed before deregulation. High redundancy, however, can actually help keep volatility in system variables effectively low, e.g., even if price of natural gas were to rise considerably tomorrow (i.e., system volatility initially increases), stored reserves and long term contracts can continue to keep the system’s (contracted) natural gas prices low. It is clear, though, that options and strategies alone do not control the volatility of the system, which explains why the HRN must be prepared for the other conditions for performance.

“Just-in-Time” Performance, Real-Time Flexibility and Adaptive Equifinality

When options and volatility are both high, “just-in-time” performance is dominant. Option variety to maintain load and generation remains high, but so now is the volatility in system variables. High market volatility may be in the form of underscheduling or rapid price fluctuations leading to unexpected strategic behavior by market parties, while higher grid volatility may be in the form of contingencies that are not covered by nomograms or as sagging transmission lines during unexpected hot weather. Whatever, this performance condition demands real-time flexibility, that is, the ability to utilize and develop different options and strategies quickly in order to balance load and generation. Operators in the control room are in constant communication with each other and others in the HRN, options are reviewed and updated continually, and informal communications are much more frequent. Flexibility in real time is the state where the operators are so focused on meeting the reliability requirement and the options to do so that more often than not they strike a match between them that is “just enough” and “just-in-time” in meeting the reliability requirement.

More formally, the state of real-time flexibility is best summed up as adaptive equifinality: There are effective alternative options, which are developed or assembled as required to meet the reliability requirement. Substitutability of options and strategies is high for “just-in-time” performance, where the increased volatile network behavior is matched by the flexibility in options and strategies for keeping performance within formal (RRN-based) reliability tolerances and bandwidths. (Remember, “to substitute” means to make a difference, and we would expect this kind of flexibility to be highly tied to operator discretion, skills and experience; more below). The increased volatility in system behavior is matched by the flexibility in the focal organization in using network options and strategies for keeping performance within the reliability tolerances.

“Just-for-Now” Performance and Maximum Potential for Deviance Amplification

When option variety is low but volatility is high, “just-for-now” performance is dominant. Options to maintain load and generation have become noticeably fewer and increasingly

insufficient to what is needed to balance load and generation. This state could result from various reasons related to the behavior of the electricity system (not just with respect to the HRN, but also the RRN and RTE). Unexpected outages can occur, load may increase to the physical limits of transmission capacity (e.g., on Path 15); and the use of some options can preclude or exhaust other options, e.g., using stored hydro capacity now rather than later. Under these performance conditions, unpredictability or uncontrollability has increased (i.e., volatility has increased), with variety of effective options and strategies diminished or less available. For example, a Stage 1 or 2 emergency has been declared by the ISO and a senior ISO official goes outside of official channels and calls his counterpart at a private generator, who agrees to keep the unit online, “just for now.”

More formally, “just-for-now” performance is a state best summed up as one of maximum potential for “deviance amplification:” Even small deviations in elements of the market, technology or other factors in the system can ramify widely throughout the system (Maruyama, 1968). Marginal changes can have maximum impact in threatening the reliability requirement, i.e., the loss of a low-megawatt generator can tip the system into blackouts. From the standpoint of reliability, this state is untenable over time. Here people have no delusions that they are in control. They understand how vulnerable the network is, how limited the options are and precarious the balance, they are keeping communications lines open to monitor the state of the network, and they are busily engaged in developing options and strategies to move out of this state. They are not panicking and, indeed, by prior design, they still retain the crucial option to reconfigure the electricity system itself, by declaring a Stage 3.

“Just-this-Way” Performance, Crisis Management and Zero Equifinality

“Just-this-way” performance is dominant when options variety is low and when system volatility must be made low. This performance condition occurs in the California electricity system as a short-term “emergency” solution. In an electricity crisis, the option is to tamp down volatility directly with the hammer of crisis controls and forced network reconfigurations. The ultimate instrument of crisis management strategy is acknowledged to be the Stage 3 declaration, which requires interruption of firm load in order to bring back the balance of load and generation from the brink reached in “just-for-now” performance.

More formally, “just-this-way” performance is a state best summed up as one of zero equifinality: Whatever flexibility could be squeezed through the remaining option and strategies is forgone on behalf of maximum control of a single system variable, in this case load. The Stage 3 declaration has become both a necessary and sufficient condition for balancing load and generation, in this case reducing load directly. This contrasts significantly with the other three performance conditions. There the options and strategies are sufficient, without being necessary. Many ways exist to skin the cat. Only in “just-this-way” performance is the “remaining option” both sufficient and necessary. You are left with only one way, or no way.

One aspect of Figure 2-3 should be noted before we turn to shifts between these performance conditions: The two dimensions are meant to capture the descriptions interviewees have given us. Our interviewees continually distinguished between the resources they have at hand and the unpredictability and uncontrollability of events to which they are responding with these

resources. To illustrate how in the ISO these dimensions are experienced, Figure 2-4 gives some salient statements for each of the four performance conditions.

A crucial aspect of the framework distinctions made is that operating under shortages is not the same as operating under high volatility. As the operator quoted in the lower right cell puts it, when shortages are predicted, the day is “clean and simple; you go through the steps and do controlled blackouts.” The operative words here are predicted and controlled. The opposite case also holds: A day may be highly volatile, for example because of unrealistic electricity schedules, even when resources—reserves, transmission lines and congestion management—remain ample or satisfactory.

		System Volatility	
		High	Low
Network Option Variety	High	<ul style="list-style-type: none"> • “A bad day is no bids and SCs don’t follow instructions.” • “A bad day is when you start out fine and then someone cuts you 1000MW and then you lose four relays, when you go home, you’ve been put through the wringer. How do you solve this unpredictability? You don’t.” 	<ul style="list-style-type: none"> • “A good day is when there is a good set of bids and everybody follows instructions.” • “A good day is everybody submits balanced schedules.”
	Low	<ul style="list-style-type: none"> • “A bad day is you don’t get enough energy or don’t get the reserves and regulation.” • “A bad day is when markets are totally screwed up and so are the tools.” 	<ul style="list-style-type: none"> • “A bad day can be a good day when it is predicted and it is clean and simple; you go through the steps and do controlled blackouts.”

Figure 2-4
Illustration of Operator Experiences of Performance Conditions

That said, there are occasions where it is not really possible to distinguish between changes in volatility and changes in options. These are at or near the corner-points of Figures 2-3 or 2-4, where the two dimensions meet, particularly on the left-side of the framework. To begin with the upper left-hand corner of "just-in-time" performance: Is a generator not following the ADS dispatch order under conditions of high options and volatility an example of a decline in resources or an increase in volatility? Similarly, is the lower left-hand corner of "just-for-now" performance untenable precisely because having exhausted all options but the Stage 3

declaration indicates both "low options" and "high volatility" at the same time? In the left-side corner-points the electricity system is both tightly coupled and maximally complex, features of a large technical system which NAT describes as its most problematic from the standpoint of reliability management. It is these corner-points, where the risk and dangers of errors arising because of misjudgment (important in "just-in-time" performance) and exhausting options without room to maneuver (important in "just-for-now" performance) are their greatest or most pressing. Table 2-3 summarizes these and other aspects of the four performance modes, some of which will be discussed in a moment as well as in subsequent chapters.

**Table 2-3
Performance Modes Summarized**

	Performance mode			
	Just-in-case	Just-in-time	Just-for-now	Just-this-way
Volatility	low	high	high	low
Option variety	high	high	low	low
Principal Feature	high redundancy	real-time flexibility	maximum potential for deviance amplification	command & control
Equifinality	maximum equifinality	adaptive equifinality	low equifinality	zero equifinality
Operational risks	risk of inattention & complacency	risk of misjudgment because of time and system constraints	risk of depleting options & lack of maneuverability (most untenable mode)	risk of not controlling what needs to be controlled
Variables of interest	structural variables (e.g., operating reserves)	concatenating variables (e.g., cascading accidents)	triggering variables (e.g., error amplification)	control variables (e.g., shedding load for reliability requirement)
Information strategy	vigilant watchfulness	keeping the bubble	firefighting	compliance monitoring
Lateral communication	little lateral communication during routine operations	rich, lateral communication for complex system operations in real-time	lateral communication around focused issues and events	little lateral communication during fixed protocol (close to command & control)
Rules and procedures	performing according to wide-ranging set of established rules and procedures	performing in and outside analysis; many situations not covered by procedures	performing reactively, waiting for something to happen, i.e., "I'm all tapped out"	performing very specific set of detailed procedures
Orientation toward ACE	having control	keeping control	losing control	forcing command & control

Transitions Between the States of Performance

Adaptation to and coping with transitions between the four states of performance are also part of the high reliability behavior of the ISO as the focal organization in the HRN. The more states of performance required of the network and its focal organization in order to be reliable, the more reliable it actually is when it can make transitions between these states, other things being equal. We have data on certain of these moves, which are detailed in Chapters 3-9 and summarized by the arrows in Figure 2-5.

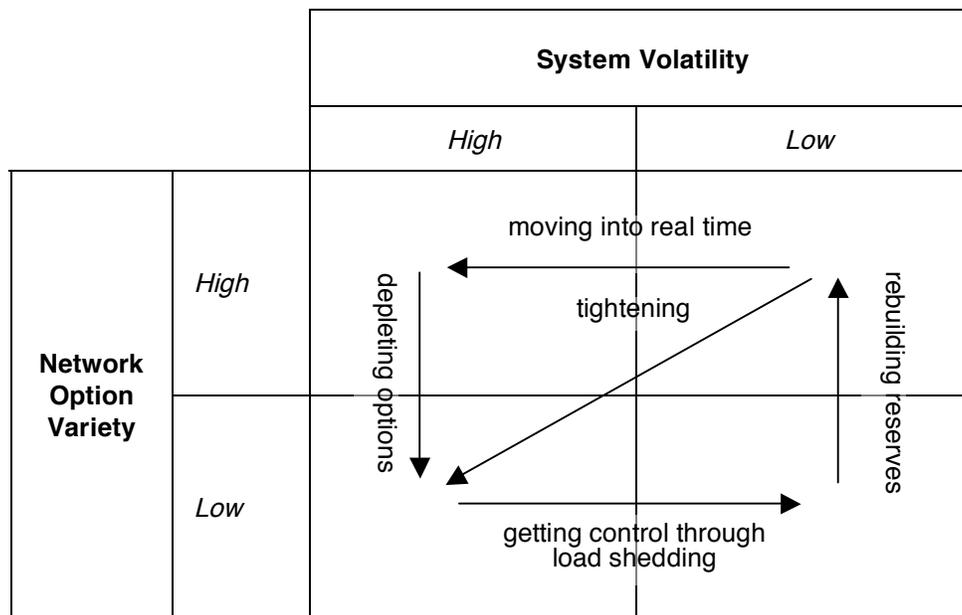


Figure 2-5
Recorded Transitions Between Performance Conditions

Starting from “just-in-case” performance, the ISO finds itself being increasingly pushed and pulled into real time (that is, into “just-in-time”) when conditions became more volatile. Operators told us about days that started with major portions of the load still not scheduled and with the predictability of operations significantly diminished. Reliability becomes heavily dependent on the ISO’s ability to pull resources and the balance together at the last minute. Because of the time pressure this brings with it, operators cannot rely completely on their highly specialized tasks and procedures, but initiate a great deal of lateral communication to quickly and constantly relay and adapt all kinds of information in order to “keep the bubble” with respect to the variables that need to be managed given the performance conditions they face. They not only have to respond quickly to volatile events, but have to make sure that their responses are based on understanding variables so that these responses do not exacerbate the balance problem, especially as confusion over what is actually happening can be intense at these times. This results in high demand for information and comprehension and a great deal of energy is spent on keeping the bubble, on tracking concatenating variables and working to keep the system, at least as it pertains to balancing load and generation. It is no longer possible to separate beforehand important and unimportant information. People from the wraparound are pulled into operations

and real time to extend the capability to process information quickly and integrate it into an overall understanding of the system. All these are typical characteristics for “just-in-time” performance.

The ISO can be pushed from “just-in-time” to “just-for-now” performance conditions when the system has diminishing margins and the options with which to keep the service and grid reliability are depleting. With decreasing options and high volatility, many potential events can now threaten control and reliability of the grid directly. Attention now shifts to identifying those changes in variables that could trigger cascading events. Rules and procedures are still important, but can also be suspended under the pressure to do whatever needs to be done to avoid blackouts or worse. As options run out or are exhausted (the ISO has tried the emergency warning, Stage 1 and Stage 2 declarations), formal and informal communication becomes less rich and more focused on direct threats.

When options continue to diminish, the ISO faces the choice to invoke Stage 3 and set in motion the procedure for controlled blackouts. Declaring Stage 3 is to move from “just-for-now” to “just-this-way” performance. Attention shifts to control variables, most notably the reduction of load directly. Communication follows this pattern and now revolves more around compliance monitoring, checking whether the procedure is carried out appropriately and whether it has the desired results. “Just-this-way” performance and its associated procedures allow the ISO to assert some command and control over a number of key variables, firm load being the most important. After a Stage 3, operator attention refocuses to expanding reserves and margins, ideally shifting network conditions to “just-in-case” performance

We also found descriptions of a “tight day,” where the ISO is pushed from their “just-in-case” performance more directly to “just-for-now” conditions. A tight day is when the forecasts indicate that the margins are going to be low; so low in fact, that all kinds of “normal” unpredictable events now result in high volatility of the system—where with high margins these unpredictability's would not have been such direct threats to reliability.

Transitions between states are most often involuntary on the part of the focal organization. An exogenous increase in system volatility pushes, for example, the ISO into “just-in-time” performance. As a result, the must respond in new ways in order to balance load and generation. What is crucial here is the ability of the focal organization to shift performance modes in response to different performance conditions outside of its control. That said, the ability of the focal organization to adapt to changing performance conditions, whether autonomously or in concert with others in the HRN and RRN, is also key to balancing load and generation. Some maneuverability again is in the form of the option to declare a staged emergency. In other words, it is important to distinguish transitions between states, adapting to or promoting those transitions, and managing performance within any one state, when it comes to characterizing the service and grid reliability in the HRN under real-time conditions. In case it needs saying, the ISO can affect the longer-term conditions independently through longer-ranged strategies to increase the variety of options (e.g., expand reserves through new generation) and/or dampen system volatility (e.g., deal with bottlenecks along critical transmission line paths).

The above framework puts us now in a position to answer the question that headed the earlier section: How does the HRN actually produce, transmit and deliver reliable electricity within the California electricity system under varying performance conditions? The answer: By having the focal network organization—the ISO—able to (1) maintain the balance of load and generation within any one of the four states of options variety and system volatility ; (2) adapt to external-generated shifts between the four states so as to maintain the balance of load and generation; (3) move out of “just-for-now” performance mode as soon as possible into “just-in-time,” “just-this-way,” or “just-in-case” performance mode; and (4) to develop longer-term strategies to increase network options and/or reduce system volatility. In brief: *High reliability (both service and grid) results when performance is sustained across all four HRN conditions.* Cross-performance adaptability—being able to shift on the fly—is as important to high reliability as balancing load and generation in any one performance mode. The ability of the ISO to maintain reliability in these four interrelated ways is absolutely crucial not simply because of exogenous factors that change outside the control of any party within the HRN but also because of endogenous factors in the ISO itself. As we shall see, there are occasions where the ISO has no choice but to involuntarily increase system volatility or reduce its options in order to balance load and generation. A new software may have to be introduced across the grid in the course of one night; the ISO may be required by its own tariff to decertify one of its major suppliers.

Some readers already familiar with the California electricity system will probably have asked by this point: So, what’s new? There have always been informal networks to rely on when things get tough, there has always been real-time constraints, and there has always been “just-in-time” performance. So, what’s really new about the operational fundamentals of balancing load and generation reliably under restructuring that wasn’t the case before?

What’s new are the same things that have made “just-in-case” performance fundamentally different from the older modes of operating under high reserves. As we saw, reserves were much higher prior to deregulation than they are now, even under “just-in-case” performance. We have also seen how the proportion of “just-in-time” and “just-for-now” performance has increased during restructuring and the crisis, and we predict that proportion will remain higher in the future, even with new generation and transmission capacity coming online.

A year after our ISO control room observations, we returned for a presentation of our the initial results of our research to senior control room staff and officials. At the end of the presentation we asked them what percentage of control room time was being allocated to each performance mode a year out from the crisis and they gave us the rough breakdown summarized in Figure 2-6.

While a different set of ISO officials might have broken the percentages down differently, the shift of the bulk of control room time to the left-side of the framework seems indisputable, even after a year of "settling down." Restructuring and the crisis it induced keep pulling elements of the ISO back to start-up mode.

Finally, the informal networks we observed during our control room observations do rely on elements of the older networks but are also fundamentally different. They are *emerging* networks, compelled to exist and form precisely because so much more time is being spent on meeting the real-time reliability requirement, that is, so much more time is being spent by

operators under performance conditions of high volatility requiring “just-in-time” and “just-for-now” performance. Indeed, reliability itself is changing. Where once grid and service reliability were by and large one and the same, increased system volatility introduced by electricity restructuring has driven a wedge between the two, most visibly in “just-for-now” performance, where the challenge of reconciling the competing demands for service and grid reliability is most palpable.

		System Volatility	
		High	Low
Network Option Variety	High	Just-in-time performance ~70%	Just-in-case performance ~10%
	Low	Just-for-now performance ~20%	Just-this-way performance < 1%

Figure 2-6
Breakdown of Estimated Percent Time Spent in Each Performance Mode, June 2001 - May 2002

Thus, the untold story of the California electricity restructuring, crisis and reliability is very different than those of the three perspectives in Table 2-2. Economists can continue to believe that reliability is what consumers pay for, and is governed through market pricing. Engineers can continue to argue that reliability is technology and good systems design and managed through proven principles and protocols. Politicians can continue to maintain reliability is the availability of cheap, always-on electricity—a right that should always be there. So too the voting public can believe the California electricity crisis was about reliability at any cost. However, without recognizing and understanding the full set of service and grid reliability challenges we present in the following chapters, all of these perspectives are not only partial but dangerously misleading. We are dealing with new performance conditions and with them new reliability challenges. We now to turn to why and how this is so.

3

THE WRAPAROUND OF MARKET AND TECHNOLOGY

Introduction

Many of us think we know the costs of electricity restructuring. In fact, we do not, because these costs have not been even remotely internalized into the prices we pay for electricity. Markets for critical services may have self-regulating features, but these markets are also on permanent life support systems that all of us pay for one way or another. The costs and nature of these latter requirements are discussed in this chapter. The wraparound is core to understanding how grid and service reliability are achieved under the differing “just-in-time”, “just-for-now,” “just-this-way” and “just-in-case” performance conditions. We focus specially on the importance of the wraparound for ensuring real-time reliability under highly volatile performance conditions, that is, on the first two performance conditions. The market may have an invisible hand, but its pulse beats from a heart in intensive care.

The wraparound buffers and supports the control room. For a senior engineer in the ISO’s grid operations engineering unit, “A good day is—we act as a backstop at the ISO. . . . You get a call from the [control room] floor, ‘We lost this’ or ‘this has gone down and we are off the map’, say, involving two or three contingencies at the same time that we never modeled before. We come out on the floor to piece the thing together.” Many ISO support staff would say something along the same lines when describing what makes a good day for them. The ISO’s wraparound are these staff and inter-staff networks supporting and buffering market and grid operations around and within the ISO control room.

It takes people to manage markets. One senior ISO engineer reported that the ISO, which was originally to be staffed by 200-225 people, now has “up to 500 plus employees.” Why? Because the ISO “needed more people for market operations and information technology people. . . . What is different than expected is that the amount of people to do the market side are many more than we thought of, much larger than before.” The wraparound is so crucial that it can survive budget cuts even when larger cuts are taking place. A very senior ISO official reported to his staff that, “Yes, it’s true that we’ve added 30 new positions, mostly marketing and lawyers. But we’re asking the departments to look at their projects to see what they don’t need now. In these times, you have to be ready to respond if there’re going to be cuts.”

Informed people can and will disagree over the merits of such staff increases and individual hires. That said, wraparound costs—and large ones at that—are inevitable, if simply because California has chosen to pursue reliable electricity through deregulated markets and high reliability networks. Several factors have to be understood here.

First, control room wraparounds were always there, it is just that the restructuring-induced crisis has made them much more visible than before. “Our expectation was that we [the ISO] would be invisible to most of the public,” in the words of a senior ISO engineer, “or only visible if something went wrong, but obviously now we are at the center.” The California electricity crisis moved the ISO wraparound from the background to the foreground—the numerous market and engineering operations staff, along with other professionals, who represent the organizational infrastructure of energy markets and grid operations. They are needed to enable the market desks in the control room to exist, function and connect their outcomes to the physical properties of the technology managed in the control room.

Second, the ISO wraparound must be large because an important role of the wraparound is to buffer the grid against market volatility and buffer service reliability from the vagaries of the market. If the market were to operate solely according to its own logic, then markets would coordinate transactions. But that is never sufficient. The aggregate of market transactions produced in this manner has then to be connected to the physical reality and constraints of the grid, which requires constant attention and non-market interventions. If this is not done, the outcomes of the markets can easily undermine grid reliability, as when generation does not follow load or congestion leads to overloads (both of which are very real control room problems). Our research leaves little doubt that markets can only exist by virtue of wraparounds to support them in coordination with the market and grid operations of the HRN control rooms.

Third, the wraparounds and their expense will continue to be large. A major argument of the report is that the California HRN is persistently incomplete in its institutional design. It is this incompleteness that justifies the need for permanent, full-time staff to be there on call and deal with the (predicted and unpredictable) threats to grid and service reliability as and when they arise, but preferably before they arise. The wraparound is a permanent way to respond to permanent volatility introduced through markets. The wraparound is there because market players profit from the very asymmetrical information and price volatility that can and have threatened service and grid reliability.

In other words, market decisions concerning electricity must be linked to one another by a reliable grid, yet all the costs of ensuring grid reliability, including but not limited to those of the wraparound, are not reflected in market calculations. In fact, the wraparound needs to be dense precisely to compensate for the market volatility and the strategic behavior generated by the unbundled players in the restructured electricity system.

The simple truth of the matter is that California’s electricity market needs to be supported by an exhaustive (and exhausted) organizational and technical infrastructure. It is needed to sustain the illusion that this market acts “naturally” in the same way that the Everglades is on its own life support system of surrounding control rooms, pumps, canals, and facilities in order to sustain the illusion that there too what is going on inside is occurring “naturally.”

Making visible this infrastructure and its costs dispels one myth about electricity restructuring in California, namely, it would reduce bureaucratic costs of administration because markets would replace command and control. Since markets for critical services, such as electricity, need to be

managed, it is always an open question if it is easier to manage markets than manage regulation in the absence of markets.

The Wraparound in Action

How does the wraparound actually work? Let's look at the HRN's focal organization, the ISO, specifically. Another senior ISO official “sees everything in a time line in terms of all operations. From the four second gen dispatcher, to the ten minute BEEP, to the hour ahead, to the day ahead, to the 72 hour which is outage and scheduling, to ops engineering and market ops, which is about the year ahead, to planning which is multi-year.” But since the role of the wraparound is to assist and buffer the control room in keeping electricity an always-on service, the wraparound takes on many features that characterize the control room in real time.

The shift to real-time activities in the HRN brought about by restructuring and the crisis was measured out in the decline in importance of the day-ahead and hour-ahead market desks and the increase in importance of the real-time imbalance market for ancillary services (Appendix A, Figure A-4). The ISO wraparound of market operations, engineering operations, scheduling and outage coordination have as well shifted considerably to real-time operations in support of the control room. In the past, the bulk of activities were done in the forward planning and scheduling units. The control room was then the fallback for managing the last portion of electricity not planned for (the control room's primary responsibility for “ancillary services” says it all). With the shift to the bulk of activities being real time, forward planning and scheduling now support control room operations directly and in ways to buffer the control room from threats that could interrupt its unicentric focus on real-time reliability. “Our core responsibility is to support real time. Drop everything and support real time,” confirms a senior manager in the ISO operations engineering unit. So too in the earlier HROs. A senior grid operations engineer with PG&E describes his role in the mid-1980s: “Our job was to provide real-time understanding of the system.”

This is not to say that wraparound characteristics are indistinguishable from those of the control room. The wraparound is best thought of in two forms: the activated and the non-activated. The activation of the wraparound depends on the conditions necessitating “just-in-case,” “just-in-time,” “just-for-now” and “just-this-way” performance. The activated wraparound becomes most like the control room during “just-in-time” and “just-for now” performance when focused on operations within the hour or the hour ahead. Since the nature and level of the wraparound support and buffering vary with performance conditions, let us return to the four types of performance modes in order to describe the implications for their respective wraparound staff.

In “just-in-case” performance, the control room reaches its maximum specialization and competence. It is the least integrated in terms of the sharing of information and perspectives. That is, integration of tasks is done through procedure, as in a classic coordination-through-task design. Each operator is performing according to established, formal rules and operating procedures. Each is doing his or her own job. There is no need to share information extensively or to have an integrated, comprehensive picture of the entire HRN in order to balance load and generation. Control room operators do not need the support of wraparound specialists or people to swap places with them or back them up. In ideal terms, it's a normal day; in organizational

terms, control room operators have slack or positive redundancy. Under performance conditions of low system volatility and high network options, much of wraparound remains non-activated, though of course staff adjacent to the control room are still developing software, testing new protocols, and troubleshooting. Wraparound staff may take a break and visit control room operators, or meetings are scheduled involving staff and operators. The big risk in “just-in-case” performance is that someone is not paying attention when they should be to unexpected changes in system volatility or network options variety. “What you want to do is avoid complacency,” a senior utilities control room official said about his operators. Problems are more likely to arise, in other words, when “non-activated” becomes inactive or “deactivated.” “One of my people,” the control room official continued, “made a mistake because he was complacent, too confident in what he was doing.”

Things heat up for everyone in “just-in-time” performance. It “can go from sheer boredom to sheer panic in one phone call,” says an ISO manager of real-time scheduling told us. The pace goes up-tempo and control room behavior becomes more integrated informally. Operators share more information, pay more attention to more variables, and overhear more of what is going on around them. It becomes increasingly important during “just-in-time” performance for operators to have an overview of what is going on in the HRN with respect to service and grid reliability. Every new piece of information potentially affects everybody else in the control room when balancing load and generation. Operators do not have a way to discriminate beforehand between the important and unimportant. This leads to the need for increased, if not maximal information-sharing, or in a phrase “keeping the bubble.” The informal, non-routine and lateral, collegial relations come to the fore, stretching into the wraparound and outside the ISO control room into the HRN. Wraparound staff start showing up or are called into the control room to help out with their specialist expertise (e.g., “he knows a lot of the San Francisco grid”) or as a back-up to an overworked BEEPer, for example. Much of this is captured in the comments of a senior ISO control room official about control room behavior during a stage emergency,

The engineers are the resident experts in situations like yesterday [when the ISO had had a Stage 2 emergency]. We augment the [control room] staff giving more manpower to deal with the situation. Workload increases dramatically. Some of the desks get really active. The fishbowl [the conference room right off the ISO control room floor] also changes, it becomes the emergency operations center. We try to improve communications to the outside in addition to helping grid operations. We try to keep lines of communication open to give good info. The same happens with the peak day call for PG&E and utilities. It’s a tool for me to get instant update with their [utilities’] boss and cut through their bureaucratic staff of operations. . . . In the stages, we augment [control room] staff to necessary levels. Engineers expert in overloading and resident experts on Northern California come out onto the floor. Some guys have expertise in particular areas and it’s good to have people to make the right decisions. You need a good support staff in this business. You are not going to know everything and you need the rest to do that for you. When I get in trouble they get the engineers. They have this intuitive ability to know when they’re needed on the floor. I don’t have to call them.

Our own observations in the ISO control room also parallel the official’s. During one emergency, as the ISO moved into a Stage 1 on its way to Stage 2, the control room team began to expand. The manager of the shift managers came out and started to help the shift manager on the floor.

Wraparound staff, who had been BEEPers, also come out to help the BEEPer on duty. The WSCC security coordinator began to walk around, talking more to the shift manager and the gen dispatcher. Specialist engineers augmented the scheduling and transmission operators. Operators and others on the floor were constantly scanning the computer monitors or wall monitors looking for information, e.g., the shift manager was watching what was happening to reserves, while the gen dispatcher constantly watching the Area Control Error (ACE). As they moved into Stage 2, the ISO senior control room official came out on the floor. He knows a great deal about out-of-state generators. In the view of one informant, if there were to be deals to be made over the interties he had the authority and connections to make that work. We were told, “He also knows the system inside out. He’s. . .another special pair of eyes.” The big risk here that is driving the activation of the wraparound and expansion of the control room team under “just-in-time” performance conditions is the operator risk of making misjudgments under the pressures of time. The more information and the more expertise interpreting it, the lower the risk of operator error in such situations, other things being equal. Because of its importance, we return to the wraparound’s greater role in “just-in-time” performance below.

“Just-for-now” performance is also very fast-paced and best summed up as “firefighting.” When options become few and room for maneuverability boxed in (e.g., when load continues to rise while imports become much less assured and predictable), control operators become even more focused on the big threats to balancing load and generation. Under “just-for-now” performance conditions, both grid and service reliability are much more readily vulnerable to threats, such that virtually all that operators and wraparound staff need to know is on the ISO control room monitors: load, generation, paths. As options become depleted, wraparound staff come to have little more to add. There is less need for lateral, informal relations. Operators even walk away from their consoles and join the others in looking up at the big board on the side wall. “I’m all tapped out,” said the BEEPer and gen dispatcher on the day we were there when the ISO just escaped issuing a Stage 3 declaration (a one-hour warning, however, had been issued). Operators and support staff are waiting for new, vital information, because they are out of options for controlling the ACE themselves. Things are clear: Those with the authority have to make a decision to declare a Stage 3. Service reliability has to be given up in the name of ensuring the reliability of the physical grid. Too much load is chasing too few megawatts. The big risk here is that the right decision is taken too late, leaving operators, wraparound staff and ISO officials with little or no room to maneuver if and when controlled blackouts do not bring load down to generation fast or predictably enough.

Under “just-this-way” performance conditions, the decision to shed load has been taken and now information is centered around compliance. The vertical relations and hierarchy of the control room extends into the HRN, even to the distribution utilities in their rotating outage blackouts. Formal rules and procedure move center stage, including the declaring and ending stage declarations. Here the HRN looks most like the structure it was before restructuring, but only because of the exceptional vulnerability of grid reliability to threats. Interruptions to service become paramount. With Stage 3, operators reassume their responsibilities and roles. Wraparound staff remain ready to help and back up, if and when needed. But the bigger questions loom: Are the load interruptions bringing load back into balance with generation? Where are threats of cascades and grid islanding? When might the blackouts be lifted? The big risk here is not controlling what needs to be controlled.

Four Qualifications about the Wraparound

Before turning to an extended discussion of the real-time reliability role of the wraparound in securing “just-in-time” and “just-for-now” performance, four points are in order. Again, let's keep with the HRN's focal organization, the ISO. First, the wraparound is no hard and fast organizational category, but rather a rough mapping of different ISO support units. Market and engineer ops are in the wraparound but not in the same way, while staff in transmission operations and planning may also be drawn into the activated wraparound from time to time though not their entire units. Second, what distinguishes the wraparound and control room from each other are not just physical placement issues but cognitive, professional and power differences as well. Some of these distinctions are nicely captured in comments by a senior ISO engineer,

One thing they could do in the new facility [i.e., a proposed new ISO control room] would be to widen it so that support staff could sit in the room and be right there. The question is who should have adjacency to the control room, e.g., computer support people want to be near. This may be a status issue with the programmers, i.e., they want to be near the control room which is the heart of the ISO. Without the control room we are nothing.

One shift manager admitted that the only people he wanted in the control room were the shift crew. An ISO transmission planner described the working atmosphere of his unit. “Not the same level of stress as out there in the control room, but stressful nevertheless.” Many such comments parallel distinctions within the control room. One of the hour-ahead grid resource coordinators sketched the differences between that desk and the positions of generation dispatcher and BEEPer,

Being in the middle of the room is also very helpful. Everybody keeps an ear out constantly. . . A lot of people do that. I do it more and more. We try to be a team even though we are different departments. Being in market ops you feel I have a big responsibility to the gen dispatcher and they have a responsibility to California. But sometimes you feel separated. These guys work real time, where on my desk, it's more of laid back. . . . We are a team in here, even though we are from different departments.

Notwithstanding the preceding comment, there is, third, an important sense in which the two terms, wraparound and control room, should not be thought of as two organizational entities, distinct from each other. In a significant sense, the control room and the *activated* wraparound share the same matrix structure. The term, matrix, is used in organizational analysis to indicate a special organizational configuration, typically drawn between product and functional divisions in a firm. The product may be corn syrup while the functional units are engineering, operations and sales. Staff from the latter units may be organized into a cross-functional team to produce and sell the corn syrup, even though each functional group has its own lines of authority in the firm. So too does the ISO have a similar structure when it comes to the control room and activated wraparound during real-time performance conditions. A senior ISO scheduling manager, for example, describes how his staff is managed in the control room. The “shift manager has functional supervision over them as distinct from administrative supervision. Functionally I am over them also, but in a real-time environment the shift manager is in functional control.”

Note the importance of real time here. “Just-in-case” and “just-this-way” performance modes are more formal, official and hierarchical; during “just-for-now” and most especially “just-in-time” conditions, authority patterns become flexible and lateral, in ways much as described for HROs (see Appendix B). Indeed, the best way to distinguish the difference between “just-in-time” and “just-this-way” performance is that the matrix shifts to lateral in the former and vertical in the latter.

It cannot be assumed that “going lateral” or “teamwork” is everywhere figured the same in the control room or wraparound. As the grid resource coordinator of the hour-ahead desk indicated above, different departments are represented in the control room. A WSCC security coordinator takes up the point:

In the vertically integrated utilities, organization followed the grid operations. Distribution, transmission, scheduling and generation were integrated, and so there was one person in charge over all four. In the current system, there are market, scheduling, transmission and others, who are each responsible to different directors. Under the current system, then, they are not integrated in the control room like they were in the former utilities. This means that, operating under different bosses, each one in the control room will have slightly different and inconsistent goals with the others. The question then, is how do you integrate these different positions in the control room so that they more match the kind of organizational integration of the former utilities? I don't know. It's the control room crews that more or less have to try to recreate the former vertical integration. Teams have to do it by reporting to a shift manager that is in a similar position as the one in charge of the older integration.

The matrix structure of the control room and activated wraparound has an important technological dimension. For instance, the Automatic Generation Control (AGC) enables the gen dispatcher to go “lateral” in the HRN's technology matrix, so that the gen dispatcher controls individual generators directly (at times manually) under contractual arrangement. So too for Reliability Must-Run (RMR) arrangements administered by the Alhambra on behalf of the ISO at Folsom. “The dispatcher in Alhambra does the scheduling of specific RMR units. [They're] a small percentage of units in the control area, but crucial to reliability,” underscored an ISO control room shift manager.

Our fourth and last point is this: While the ISO is the focal organization, its *raison d'être* is networked reliability. Networks of people from other control areas, generators, trading floors, and utility control rooms are on call as well. The ISO's wraparound is not the only network of expertise and personnel drawn upon by the ISO's control room during real-time performance conditions. The informal, non-routine and personalized communications, accountability and trustworthiness of these wider networks are also instrumental for service and grid reliability. In organizational terms, these wider networks come to look very much like informal teams, having a quasi-matrix structure in high volatility situations when network resources are being depleted quickly.

The Wraparound in “Just-in-Time” Performance

The activation of the wraparound and the expansion of the control room team under “just-in-time” performance conditions is largely one directed to support and error reduction, both to decrease the heightened risk of operator misjudgment. We observed three kinds of overlapping knowledge bases activated in these situations. First is representational knowledge, such as what operators see on their monitors. The second is experiential knowledge that operators have gained from on-the-job learning as to how to operate in *these* kinds of circumstances with *that* kind of information. Third is the related but specific institutional knowledge of who and what can or cannot be trusted to do what, where, when and how. For example, institutional knowledge comes into the control room when the fishbowl starts filling up with the emergency response team and other people. “People just showing up” has also been observed in air traffic control centers during emergencies. People at these centers have many of the same functions as air traffic controllers themselves and knew what was happening by what they were seeing on their monitors. It can almost be anyone who shows up, because a great many people have the same kinds of air traffic control skills and knowledge

Part of the knowledge base is knowing when to stay out of the way. We observed in the ISO control room ISO staff letting the activated wraparound and control room operators get on with the business of self-organizing, i.e., the matrix went lateral. Officials stayed out of the way as operators and support staff got on with what they needed to do and did not manage that process. What supervision there is becomes “coordination.” “During that time,” reports a shift manager about peak days, “a lot has changed in the control room. A lot of people come through. There is a lot of activity out there. There is a lot to be coordinated.” In the same way, meetings and conference calls in the fishbowl involving the emergency response team seemed more coordinated and self-organized than supervised or designed. To behave this way means those involved have a fairly homogenous cognitive understanding of the way all of “it” works, a good example of “keeping the bubble.” Thus, if and when they fail, they fail together as a team; if and when they succeed, it is because they were all doing their jobs.

So far, we have principally focused on the support role of the wraparound but have said little about its buffering function. One sees this role clearest in the emergency response team. “We [the emergency response team] support real time in grid-ops. We shield them from media and questions,” according to an ISO operations support staff person. PG&E also has an emergency response team consisting of, among others, planners, operations engineers, maintenance people, and a customer service representative. A good “emergency” example of the wraparound’s buffering role is found in the week the Power Exchange (PX) died. For those in the wraparound, it was hectic, according to a senior ISO engineer, “The PX shutdown was a very ‘interesting’ week, no more schedules came in. So we had to invent whole new rules almost overnight. Bit of an experiment. . .” Yet, when asked how the demise of the PX affected the control room itself, one ISO shift manager said, “I thought, Oh my god, this will be bad, but it was easy. They put in other people to put in the schedules. For us, the PX was an SC [scheduling coordinator] basically. Now the bids come in directly.” Another shift manager agreed, saying that the “PX was nothing more than an SC...That was felt in the day ahead market. That was it—[just a] bunch of people we gave a new ID to.” One very important role of the wraparound is to protect the control room from such exogenous changes and system volatility.

In the more formal language of the Chapter 2 framework, the wraparound achieves the buffering and support through adaptive equifinality and substitutability. Wraparound staff literally substitute for control room people, just as operators or staff within the wraparound or control room can take the place of others there. Core is cross-training of operators in the control room, where for instance the gen dispatcher has been a BEEPer, or the day ahead desk coordinator a gen dispatcher. As one generation dispatcher told us,

It makes a lot easier if you cross train. . . You know what the other guy's job entails. You know the kinds of conversations back and forth. Cross training makes it easier to run the BEEP and vice versa, you learn how to take more MWs, subtle little things so you know more. It makes life generally easier.

Such substitutability extends into the wraparound and throughout the HRN. According to an ISO transmission planner, “Engineers in [the ISO’s] Engineering Ops and in Planning are almost swappable. . . . To get interaction we’ve located the groups close together. What works is mostly some informal rules. You know, ‘Don’t give me a transmission proposal unless ops has looked at it’ . . . By keeping the groups together we try to prevent gaps. We turned down the offer to move to better offices for that reason.” As for the HRN, substitutability is core to ensuring reliability, if simply in terms of trades, SCs or imports substituting for each other when it comes to balancing load and generation.

As noted in the Chapter 2 framework, substitutability is key for “just-in-time” performance, where system volatility is matched by flexibility of network responses. This view contrasts with that of Normal Accidents Theory. For NAT, the ability to substitute elements is crucial for loosely coupled systems and linearly interactive ones. Yet we find that the same substitutability is key to “just-in-time” performance, when the technology is in Perrowian terms very “tightly coupled.” According to Perrow,

What is true for buffers and redundancies is also true for substitutions of equipment, processes and personnel. Tightly coupled systems offer few occasions for such fortuitous substitutions; loosely coupled ones offer many.” (Perrow, 1999 [1984], 96).

On the contrary, our research indicates that adaptation and substitution are crucial variables in promoting reliable behavior of the tightly coupled, complexly interactive grid created by restructuring.

This is not to say that everyone in the control room or wraparound is happy with “just-in-time” adaptation and substitution. For one WSCC security coordinator, using band-aids to fix the technology did not solve things, but led only to more band-aids. “It is a system that has been constantly patched up. We are constantly improving the system with patches,” described one grid resource coordinator of the hour-ahead market software. In response to our question, How does the control room maintain reliability under such high volatility, a shift manager responded immediately, “I have six words: By-the-seat-of-our-pants.”

Conclusion

Band-aids, patch-ups and by-the-seat-of-our-pants make people nervous. Yet “just-in-time” performance is nothing new. It too was and is found in conventional HROs. What is different is the portion of operator time spent under high volatile performance conditions.

It is useful to understand “just-in-time” performance of ISO operators and the activated wraparound along two rough dimensions: the scope of their options and the context of their operations. Scope combines both the span of options available to operators and staff and the size of groups or teams involved in making these options effective. The context of operations can be high or low, where “high context” is made up of messages, meanings and expectations that contribute to self-regulating and self-organizing behavior. In high contexts, relational communication is very important, e.g., the raised eyebrow, the change in voice, the silences. Low contexts, in contrast, “decontextualize” relational communication. Communication and meanings become unambiguous references to objects and states, where any ambiguity is considered noise or miscommunications (Batteau, 2001: 201-211). In earlier terms, one can think of “relational” communication as based on both experiential and institutional knowledge, while referential communication is largely based on representational knowledge.

The resulting typology in Figure 3-1 defines four types of control room and wraparound behavior. Within the ISO control room and activated wraparound, examples of each different type of behavior are found: highly relational, case-specific team behavior (e.g., between gen dispatcher and BEEPer, or gen dispatcher and shift manager); behavior as it moves from warnings into Stage 3 emergencies (where people from the wraparound “just show up...”); or market behavior, either through discrete scheduling procedures (handful of SCs submitting in supplemental energy bids) or through market-wide transactions for ancillary services.

Figure 3-1 highlights why teams with flexible authority patterns are important in an otherwise hierarchical ISO. We would expect to find all four cells in the focal organization as it manages both grid and market that necessarily span small and large scales and high and low context situations.

		Scope of Operations	
		<i>Small</i>	<i>Large</i>
Context of Operations	<i>High (Relational)</i>	Example: Discrete team behavior	Example: Behavior for ensuring real-time reliability during stage emergencies
	<i>Low (Referential)</i>	Example: Discrete scheduling behavior in the day ahead and hour ahead markets	Example: Market transactions across the ISOs three markets

Figure 3-1
Typology of “Just-In-Time” Behavior of Control Room and Wraparound Staff

In each performance mode—”just-in-case,” “just-in-time,” “just-for-now” and “just-this-way”—one finds all four types of Figure 3-1 performance behavior. All four types of behavior are present in each performance mode. That said, the proportions and mix of types change, since different states of system volatility and options variety entail different contexts and scope of operations. We observed that relational communications increases in “just-in-time” performance relative to the referential (but both are necessarily found), where “getting direct feedback” (referential) has to be complemented by “keeping the bubble” (relational) in such highly volatile situations.

Thus, within our framework, the probability of failure increases when there is a mismatch between what the context and scope require and what is actually provided by way of communications. A senior manager in the ISOs operations engineering unit noted,

There is a certain part of the system that is in people’s heads. I know that [the procedures] are getting thicker, but I’m not sure that means we are getting smarter. The procedures represent a good deal of it [our experience and learning], not all of it. Some aspects are like folktale knowledge, difficult to quantify or put onto paper. One of the problems right now is that we have an awful lot of procedures, too many procedures. Too many for any one person to know.

When the referential displaces the relational or vice versa, but context and scope require otherwise, then risks of misjudgment increase.

Now we can see why behavior at or near the left-side corner-points of Figure 2-3 is so risky. There it is not possible to distinguish between change in options variety and change in system

volatility. The operator cannot only trust his or her representational, institutional, or experiential knowledge bases to tell them what is actually going on. Again, is the generator who does not follow an ADS dispatch order reducing the gen dispatcher's option set or increasing the volatility s/he faces? If it is both, then what is the gen dispatcher to do? Try another generator and risk getting the same response, thereby depleting options and increasing volatility even further?

Of particular concern are control room situations where the context requires relational communication and behavior that cannot be easily analyzed or replicated, but where formal procedures and protocols demand referential “information” that admits little or no discretion and flexibility. High context situations are being treated as if they were low context ones. A great deal of “just-in-time” performance is “outside analysis.” It is “making it up as you go” or “shifting on the fly” in the face of increased unpredictability or uncontrollability of events. To mandate thorough, fact-finding analysis before taking action on the basis of experience is to ask for problems under conditions of high volatility. In this way, then, the ability to patch up, band-aid, and get by on-the-seat-of-our-pants looks less considerably less risky than not being able to do that at all.

Thus, the comment of a senior ISO official is cause for concern about a future that depends on keeping relational and referential communication—and with it institutional, experiential and representational knowledge—in balance under changing conditions rather than one type dominating another for all the time. This senior official concluded his last interview with us by saying, “But I am worried about a skilled operator interface. Like how do we train the dispatchers in the control room. We are short on that. Look at the control room. [Three people there] are all retirees. I need the people with the knowledge, skills and completeness to handle the job in the future. . . Now we are using retirees!” The control room needs people with experience, but it also needs people who know what it means to operate at the different scales and through different phases created by restructuring and its in-built volatility and turbo-market conditions.

Finally, it is worth repeating the wraparound’s major design implications to which we will be returning in the Chapter 10 recommendations. First, markets are managed by people and people manage the grid. Both activities require very, very smart people indeed, who are decidedly not the cipher in economic theory who needs only to know the difference between two prices in order to act rationally. Second, these wraparounds are a permanent feature of market-based electricity systems as they were in regulated ones. It takes a great many people to handle the persistently unfinished business of operating in real time, under conditions of systemic hyper-volatility. Third, the wraparound depends on relational communications and not just referential ones when rationality itself demands adaptability between highly variable performance conditions. That is, markets are not going to be allowed to manage themselves precisely because no one can afford to let service and grid reliability be “managed” that way.

4

ELECTRICITY AS AN ALWAYS-ON SERVICE

Introduction

Almost anyone who has lived in a developing country knows that electricity is not always on, if it is there in the first place. Yet this design feature of an always-on service is core to the high reliability provision of electricity in California, and for that matter, in most of the United States and OECD countries.

Most Californians want electricity to be available 24/7. It should be there, whenever we need it. This chapter shows that the demand for always-on availability arc-welds service reliability to real time performance conditions in the California restructured electricity system. The connection between availability, reliability and the always-on service is one that depends very heavily on elements core to the ISO's wraparound. In this chapter, the focus is more on the ISO's role in the wider HRN, however. Our specific concern here is on the role and importance of relational and referential communication and feedback based on representational, experiential and institutional information for real-time reliability of the HRN as a whole and not just of any one control room.

What Does “Always-On” Mean?

The always-on feature of electricity has this rationale: Electricity is too important not to manage all the time. Indeed, the electricity critical infrastructure has been *deliberately* designed to require its full-time management, around the clock, every day. So much has been made dependent on electricity's availability that every electricity system design and re-design presuppose its full-time management, and all major players assume this management to be necessary. So much of our lives and livelihoods are stocked with services riding on the reliable and easy access to electricity that the costs of not attending to this critical infrastructure far exceed the costs of constantly having managers on their toes in ensuring that it seems to work without much of a hitch. And like other truly critical infrastructures it is entirely taken for granted by many segments of the population.

We Californians are like the arid-land pastoralists and nomads—they too are accused of “overstocking”—who keep large herds as a way of increasing their dependency on them and thus the need always to be managing herds, given the massive societal costs associated if they did not. Our deliberate and utter dependency on always-on management is the overhead we pay for our respective high reliability systems. A senior engineer in PG&E's grid operations describes some of the high stakes in having to be always ready when electricity has to be always on,

‘Is the worse thing you could do not to act fast enough?’ Yes. It’s like when you need a shot of atropine in the heart. Here you have 3200MW, one line goes out, and you need to have an alternative immediately. . .[B]ecause of the privilege of operating the system at high reliability or stability we can’t wait minutes to act. . .You have to react right away, in seconds. So it means tripping generation, as you can trip hydro quickly and fairly easily get it back on line. The RAS [automated Remedial Action Scheme] means that you can have a 4800MW interconnection, where without RAS it would mean a 1000MW limit.

Just what is “always-on?” Surf the web and you will see definitions such as this: Your system is always on the Internet, 24 hours a day, with a reliable and safe connection running at speeds many times faster than any dial up connection can provide. It is so easy and convenient in the same way you access electricity or water...and so on. Sounds a bit like those “perpetual care” contracts, but patently a significant part of the reliability of electricity is that it is always there, all the time, like a dedicated internet line. As with all such commitments, this one requires the attention of managers, including their persistent alertness and readiness.

Behind the experience of electricity as always on lies a core reliability paradox amplified and reinforced by restructuring. Always-on management of electricity as an always-on service may actually require turning that service off in order to keep the grid always reliable. This trade-off between service and grid reliability is one that is most visible under high volatility conditions. In the past those conditions have come about because of storms, fires and grid cascades. Now, of course, price and market volatility have made those conditions much more likely for ISO operators. This permanent cleaving of service and grid reliability under restructuring surfaces in many ways, as subsequent chapters show.

Restructuring has changed the always-on character of electricity in other ways. For many generation operators, reliability in the older integrated utilities was equivalent to the constant availability of the plant’s generation. One plant operator told us, “Availability and reliability are one and the same thing. The objective is to take care of things, to anticipate problems, so that you are always available.” Indeed, the goal of the other regulated utilities was to have the generation available as close to a 100% of the time as possible.

Restructuring changed that. A senior generation executive of one of the state’s energy suppliers described the change,

In the old days, reliability was being able to ensure your plant was 90% available or 95% available. Now availability means nothing if there is no demand for it. Having a unit available when it wasn’t needed was ridiculous. It meant keeping the unit online when it should have been unavailable in order to meet the right level of maintenance required. . .Reliability [today] is hard to measure on an annual basis. . . I was once asked if I can get 96% availability on my units here. I said, you don’t want me to get that. It’s not a good business decision for us or for the customers, to be available 96% of the time.

From this perspective, keeping a unit online and available at any time has to be traded-off against the effects on its availability at other times, when it may be commercially more important—i.e., prices are higher. The calculus of that trade-off includes, for example, taking penalties for not

meeting prior commitments or doing maintenance now when commercially it is more important to be sure to be available at a later time. A number like “95% availability” does not reflect these tradeoffs at all. The remaining 5% may be a lot more important than large parts of the 95%. “What you really want to do,” the generation executive continued, “is minimize lost opportunities and penalties. . . . Success is measured in terms of [not of the older notion of availability but] ‘commercial availability,’ [that is,] produce when market prices are above your cost of produce and minimize penalties by meeting your commitments.”

The consequences of such behavior by private generators and SCs within the HRN have been to increase the system volatility that the ISO has faced. In fact, a defining feature of the California electricity crisis has been such behavior on the part of generators necessitating the ISO’s shift to real-time markets and last-minute out-of-market transactions in order to ensure load and generation are balanced.

The ISO has sought to keep electricity always available, no longer through dedicated generators but through increasingly last-minute transactions connecting independent generators through the distribution utilities within the HRN. In terms of our ISO reliability framework, the restructured variety of “always-on electricity” cannot be satisfied by any one “optimum” performance condition or performance mode, be it “just-in-case,” “just-in-time,” “just-for-now” or “just-this way.” Because of restructuring, there is no optimum. Rather always-on electricity requires the ability of the focal organization to shift on the fly and adopt different performance modes as conditions warrant—which they do much more often than before. In some cases, this means that to keep the grid reliable the ISO must sacrifice service. What we have been calling the “cross-performance adaptability” of the focal organization (Chapter 2) can be termed “always-on management.”

This chapter focuses in more detail on this challenge of adaptability and on one of its most important requirements, namely the demand that feedback on reliability be immediate, continuous and unambiguous. Immediate means the feedback is real-time information, continuous means the feedback is always there 24/7, and unambiguous indicates that the feedback provides patterns that can be read without mistake or without incurring highly consequential errors. We have found such feedback to be associated with informal processes of accountability and trustworthiness, processes of training and recruitments, and shared communication in the HRN.

Always-On Management as Cross-Performance Adaptability

What do these shifts between performance conditions actually look like and mean for control room and HRN-wide behavior?

We can analyze the shifts in two ways: specifically in changes in individual behavior as performance conditions change and, more generally, in terms of the differences in performance conditions themselves. An ISO shift manager described how what we have been calling “just-in-case” performance differed in terms of tasks undertaken compared to “just-in-time” and “just-for-now” performance. Table 4-1 shows the percentage time breakdown the shift manager gave for the various activities this manager undertakes during “normal” and “peak” days:

**Table 4-1
Percentage Time Breakdown of Shift Manager Tasks**

Normal day: Activities and time spent	Peak day: Activities and time spent
<ul style="list-style-type: none"> • Routine 30% • Procedure comments/reviewing 20% • Coordination on the floor 20% • Meetings 15% • Normal dept. communications 10% • Reviewing outages 5% 	<ul style="list-style-type: none"> • Coordination on floor 30% • Firefighting 25% • Dispersing info to departments 15% • Conference call (plus prep time) 10% • Issue operating orders 10% • Routine and logging 10%

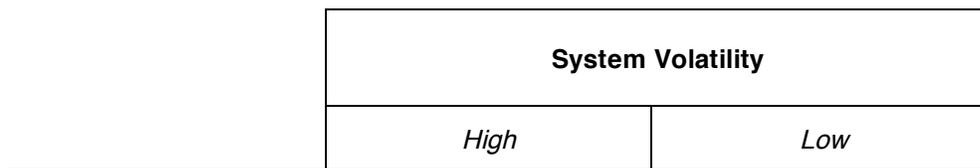
As the shift manager moves from normal to peak conditions, coordination requirements increase, while “firefighting” emerges as its own noteworthy requirement. Such shifts are common for the other operators in the ISO control room. A beeper describes normal and peak day activities as follows:

My calls are with the SCs, maybe three or four calls on a slow day, but on days when we’re busy, the phone rings off the hook. On the whole how many people do you have calling you? It’s hard to say, sometimes 10 people calling in an hour; on in a slow day it is 3 or some calls. . . When you get to know them over time, there’s some that have more expertise, and there’re others less skilled.

Our interviews describe much the same increased tempo and pace in activities, while new ones are added as operators shift from "just-in-case" performance to the other types of performance modes.

Not only do some tasks change from performance mode to performance mode, but the nature of the moves themselves change. Part of always-on management is that control room operators always want a way out of any situation in which they find themselves. If they cannot avoid being pushed into a position of low options variety and high system volatility, they want to move out of it quickly, even at the cost of a potential sacrifice of service reliability.

Cross-performance adaptability and the major differences in the way always-on management is undertaken within the HRN by the focal organization are summarized in Figure 4-1.



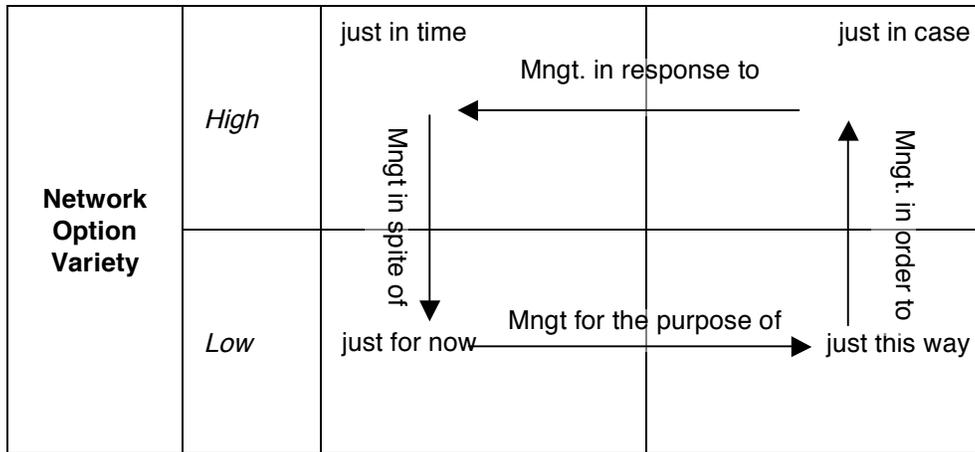


Figure 4-1
Four Types of Always-On ISO Management Under Real-Time Performance Conditions

Managing in “response to,” “in spite of,” “for the purposes of,” and “in order to” entail different ways the operators are pushed and pulled into and out of performance conditions and across which they can more or less adapt.

Let us turn to each. When system volatility increases while HRN options variety remains comparatively high, ISO network management becomes one of “managing in response to these changes” so as to keep electricity always on. When volatility remains high, but network options are being depleted or exhausted, ISO management becomes one of “managing in spite of fewer and fewer options and high volatility” so as to keep electricity always on. A Stage 3 is declared “for the purposes of” reducing system load to meet available generation, where now what is “always-on” is not the service but grid reliability. Once load and generation have been (re)balanced as a result of the stage emergencies, efforts begin to increase operating reserves and other positive redundancies in order to get back to “just-in-case” performance and thereby reduce the vulnerability of the HRN to surprise. The first two forms of always-on management are a consequence of system-shift conditions, while the second two are directed toward system recovery and restoration conditions. Of course, each type of always-on management has its own reactive and proactive components, and there are other shifts (along the diagonals or moving clockwise), which are an important part of the cross-performance adaptability of the HRN’s focal organization. Whatever the case, this adaptability—or what we call here “always-on management”—is core to the focal organization’s service and grid reliability mandates for ensuring the always-on service.

Feedback on Performance

A major effect of California's electricity restructuring and crisis has been to make always-on management more problematic. In response, “networks” based on feedback and pattern recognition within and across the control rooms and wraparounds of the HRN have become absolutely crucial in meeting the reliability requirement of balancing load and generation. Pattern recognition is important because the urgency of real time makes it crucial to “read” feedback in terms of signature events that can reliably guide the balancing of load and generation, in the

absence of operators having to have full causal knowledge of the system in the process. This substitutability of reliability-enhancing signature events for complete causal understanding of the system is particularly important, as up to this point the major approaches to large technical system reliability assume full causal understanding is necessary for reliability, while its absence increases the risk of normal accidents.

The feedback we observed was most useful when it was immediate, continuous and relatively unambiguous to parties within the HRN, particularly when undertaking “just-in-time,” “just-for-now” and “just-this-way” performance. For example, senior PG&E control room officials are always on call. When there is an emergency, their pagers go off, their cell-phones are called, their land-lines phones and voicemails are contacted, and a continuous stream of messages are left until the one called answers or responds. The messages are immediate, and they are, singly and in aggregate, relatively unambiguous with respect to the need for action.

Feedback flows throughout the control room and wraparound, all the day. One ISO shift manager illustrates:

The weather service that we’ve contracted, Weather Bank, . . . will call twice a day at 6:30 and 13:00. I’ll look at the temperatures, potential for high winds, fog precipitation. They are very thorough, so I don’t usually have to ask them anything. I look at the scheduled outages for tomorrow and look at the load the day before. If they predict fog, for example, in the Bay Area, I focus on that and try to get clues for loads with fog. With heavy fog and dirty lines you might even have problems to operate and mess things up. . . . Then, we have a nine o’clock meeting with critical staff in the fishbowl. This meeting takes place every weekday. Usually the director of grid operations is there. We’ll have representatives from market ops. They tell us what the load forecasts for the next days are and other issues in market operations. . . . It’s really an open form of communication of all departments closely related to the floor. This is definitely more real time focused. . .

This shift manager goes on to explain how the feedback and information demands vary when the day is a peak one rather than a normal one (see also Table 4-1):

A high load day differs from a normal day. . . If I haven’t worked the day shift the day before, I have to get into things fast. Looking at the load forecast, state of the transmission lines, the internal generation status [i.e., generation present in the ISO control area as opposed to imports from other control areas via the interties]. How are the other areas doing? Can we buy out of state? At 7:30 [a.m.] there is the peak day meeting with a lot of people. These meetings give you a lot of information about the last day, how much load was used, how much load may have been interrupted, the status of transmission and generation. Then the meeting starts discussing the interties and the forecasting of all that for the current day. For example, talk about a major unit that is supposed to come off- or online. It is a real conference call with a lot of people. The utilities (PG&E, SCE, SDGE [San Diego Gas and Electric]) and all the municipalities, some of the public agents (CPUC, CEC), and other government agencies that have a need-to-know. There is a representative from WSCC who calls in. The director of grid operations usually directs the meeting. The purpose of the meeting is to keep everyone informed. Depending on the type of day, we usually have more of those meetings. There is one at 10:00/10:30 and 13:00/13:30, basically to keep everyone informed. . . . When our

load forecast is off, the utilities like PG&E have their own load forecasts and I can compare these and discuss with them about significant differences.

While feedback should always be on, it may not always be right on. In getting to know the ISO control room operations, we were surprised by how many indicators and measures were unused or ignored: the huge wall panel showing the grid which is rarely used by operators, as every thing is of importance is on the computer screen or sidewall monitor; the ACE showing up as red, when in fact it had long moved to within the reliability bandwidths; a ticker-tap monitor that was for the media but of no importance to operators; an hour ahead desk that is really three hours ahead; alarms which are not accurate; missing telemetry information on generators that are really “there;” and direct dial phone numbers long out of date.

Not surprisingly, then, the feedback that is most useful to operators is real time and from reliable multiple sources. When feedback is not so conformed, many problems can arise. First, the real-time issue. “Day-ahead schedules do not reflect reality,” said a lead official in ISO marketing operations, adding “[they] should be closer to what it really is in real time.” A senior ISO engineer expressed a related problem: When feedback is not real time, no one knows how operators are actually doing:

[Another] missing piece is that the dispatcher doesn’t actually see the results of his mistake. We should put up a board that says, “this is how much we paid in fines last week” [because of such mistakes]. Right now, there’s no feedback loop for this. Now it just depends on management style [of the shift manager and control room operator concerned], and I think we need to make that loop stronger. Plus I would do more post-mortems [after emergencies], but that doesn’t happen very much, at least as far as I’ve seen. If you want to improve performance, you need feedback.

The lack of performance feedback is singled out by interviewees as a problem in responding to the California electricity crisis: “One of the biggest failures is that there weren’t any feedback processes, they designed things without that feedback to make them better,” in the words of a senior grid operations engineer at PG&E. The virtue of real-time information is that it is more trustworthy and salient to control room operators than delayed or lagged information. According to a senior generation executive, “our job [in the plant] is specifically to tell the trading organization what we can actually meet in terms of commitment. We also tell them this in real time. . .”

Not only should the information be real time, better yet it should be from multiple, reliable channels and sources. “Based on previous days, you can read logs to see what you need to be aware of . . . They had a problem on that day like the one I’m having now and I want to verify it, see what they did, so I use all resources to be prepared,” we were told by a PG&E shift supervisor in their Transmission Operations Center, “There are all these tools, we’re monitoring all that, we’re getting feedback from the field, and we’re all getting different information, we all want to be aware of what everyone else is doing in this room.” To put it differently, the real-time reliability of “just-in-time” and “just-for-now” performance relies on signature events rather than full causal knowledge, but this works best when everybody in the control room and activated wraparounds are paying attention to what could go wrong in doing so. “Ten people looking at

[the same urgent problem] with different experiences means you can get a better grasp for what the problem is,” concluded the senior generation executive.

The differences in feedback and their timeliness are summarized more formally in Table 4-2 in terms of the threats and errors identified by the feedback which affect the ISO’s balancing of load and generation.

Table 4-2
Feedback and Discovery of Error in the ISO

		Feedback	
		<i>Continuous</i>	<i>Discontinuous</i>
Discovery of Error	<i>Immediate</i>	“Always on” CPS2, ACE, ADS, AGC; real-time price information on trading floor	RAS; SLIC for scheduling and logging
	<i>Lagged</i>	CPS1 violations and settlements	Post-mortems after crises

Measures, ranging from real-time prices for energy traders to CPS2 violations, are immediately obvious to control room operators. “Traders are looking at the screen all the time,” reports a senior generation executive. The automatic dispatch system (ADS), for example, gives a reply from generators within minutes: “ADS has saved so much effort and time for us” and “ADS has made the job of these guys a lot better,” reported a gen dispatcher and day-ahead grid resource coordinator respectively. “Twelve to 14 months ago ADS came and made the transmission job a lot more bearable,” summed up an ISO control room shift manager.

Yet sometimes the ADS never immediate enough in an environment that places a premium on multiple channels of confirmatory information. “We still make the phone call, no matter that you can send dispatch electronically. You call first, check whether they are able to do it, then actually send the dispatch to them,” reported another shift manager. So too do control room operators manually manipulate generators through AGC in real time. Hands-on confirmation is as real time and unambiguous as feedback can get from an operator's perspective.

Other measures are definitely not immediate. Returning to Figure 4-2, CPS1 violations, while based on continuous data and feedback, are only calculable at the end of the day and settlements

much later than that. Activation of a RAS is only periodic, while the SLIC system for scheduling and logging is updated only as and when needed.

The timely admission of control room misjudgments is essential. We observed a BEEPer making a mistake that was obvious to those around the beeper and readily admitting it. In reply to our question why the ACE was at one point continuing in a CPS2 violation, a gen dispatcher said, I miscalculated. When a shift manager asked how things were going, another gen dispatcher admitted to having erred earlier in the morning which caused a great deal of overgeneration with which they had had to cope. Neither gen dispatcher defended the mistake. At an ISO conference meeting, a very senior ISO official reiterated that “in our [ISO] culture, if a number is bad, we try to find out why, instead of blaming each other as [some other state government officials do].”

The ready and open admission of error (which is also an important feature of HROs) occurs in a setting where feedback implies accountability for acting upon the information provided. As an example, the new Energy Management System (EMS) system in the ISO control room will give operators and their wraparound greater information which should help them catch big problems or mistakes before they happen in the system. A senior official in the ISO’s grid operations told us,

We have automated programs [that] look for problems. . . We want and are going to get an EMS system that allows you to take a real-time snapshot of the system that you can download and model with offline. . . . The big thing is that the new system information and contingency analyses would be available on the floor, for use in real time. . . . What the contingency screening would do is that they, the operators, would get a lot of real-time violations that they didn’t know about before and thus will call us more. They are now in some respects flying blind on this stuff. Now they would be seeing more and going deeper into the system. Yeah, going deeper is true, but we have been on the hook for the grid reliability all along so that’s nothing new.

The ability to identify big mistakes before they happen and admit error after mistakes reflect an important fact of life for the operator: Continually prompt and clear feedback over a wide range of performance factors crucial to balancing load and generation in real time is what keeps people and their information as honest and useful as they are.

Let us now turn to the extent to which and ways in which feedback “keeps the system honest.”

Keeping the Network Honest and Accountable

The honesty is glimpsed in the informal communications, trust and accountability with which parties in the HRN deal with surprises, unpredictability's and contingencies that were not originally anticipated or even allowed for officially. Operators come to learn and know who is screwing up or gaming, or is reliable in the face of the expected and unexpected. Such institutional and experiential knowledge becomes second-nature. “You don’t have to look at the trend lines, you know what to get. . . sort of intuitively,” said one of the ISO hour-ahead grid resource coordinators.

Even though a culture of reliability does not exist in the network-wide provision of electricity, official and unofficial systems of shared knowledge, trust and accountability for actions and their consequences help compensate for some information asymmetries, strategic behavior and incompleteness in the HRN. Reliability networks necessarily involve divergent interests, and trust within the network can be defined as divergent interests without polarization (without open cleavages). Distrust, in turn, is the state of divergent interests within a network having become polarized. Accountability in such a network is holding these divergent interests, whether trusted or distrusted, responsible for the balancing of load and generation.

Key to trust and accountability is reciprocity, both official and unofficial. A senior ISO outage coordination official describes his unit's relationship with generators,

There is a lot of informal give and take. A good example is some of the [generation] units in the Bay Area. They are very interdependent. There is a high concentration of load, but not a high concentration of generation and transmission, so you have to do work to only allow certain outages in coordination with others. This works OK if it is the same company, but if they are owned by different companies, then we can't get the cooperation. You need authority to get them to move, and there have been times when they have moved and that cost money to them, and we reciprocate unofficially.

From elsewhere in the HRN we received similar views, making it clear that there are interactions to deal with the difficulties arising from the market design that are not captured in the tariffs. The importance of unofficial reciprocal arrangements and working relationships between the ISO and the generators, particularly with respect to outage coordination, appears differently from the perspective of some regulators. A CPUC interviewee described a source of tension between the ISO and CPUC this way,

An example of a rocky road in our dealings with the ISO is the issue of plant inspections. . . The ISO is responsible to coordinate all outages and we are responsible for evaluating the real outages that occur by asking: Was it authorized? . . . Yeah, it's sort of a classic "Who is responsible and accountable?" Both of us say "we are!". ISO says, "We need to work these generators on a day-to-day basis in a comfortable way and we don't want to be seen as enforcers. If that happens, power plants will no longer be open to us." So we say: "Fine we [the CPUC] want to be the enforcers. You just collect the information and give it to us." And then the ISO expresses the concern that "We don't want to give you the information because they [the generators] will know we are going to give it to you." The ISO wants to stay in the role of a friendly partner with the generators in making the market work and we, given our history, are more comfortable with a role as enforcer. Culturally we're both in a different place.

Many others we interviewed also underscored the importance of informal, unofficial cooperation and reciprocity. We were told that the ISO would keep the market open for some parties and under certain conditions, if this would enable them to submit better schedules. Similarly, it was reported to us on more than one occasion that information was shared within the ISO when operators needed it, even against the advice of legal staff. This out-of-channel sharing of information may look questionable, but we have no doubt that such information and collaboration was pivotal in containing volatility and/or adding options to the ISO and thus achieving the relatively high level of reliability observed in California during 2000 and 2001.

Indeed, we believe restructuring left HRN parties no option but to informally cooperate across interorganizational boundaries when service and grid reliability are their paramount objective.

By deliberately designing these large technical systems so that we have to depend on them for multiple services all the time, we have chartered them to act informally on our behalf. In this way, the always-on management mandate compels pattern recognition and the type of feedback that increase network trustworthiness, set common priorities across organizations and integrate and pattern the bubble of real-time operations. The mandate of keeping electricity an always-on service goes a long way in explaining the existence of largely invisible, informal networks that keep the HRN together, especially since these networks transgress other mandates, formal responsibilities, even laws at times. All this is achieved under competing interests, conflicting organizational objectives and differing societal values.

Factors Important for Trust and Accountability

A key factor in ensuring the full-time attention necessary for always-on management of service and grid reliability is recruitment, training and retention in the ISO and HRN. The ISO as a focal organization has acted very much like the earlier HROs in giving special care and focus to recruitment, training and retention as a way of inculcating and sustaining high reliability. Pride and ownership of results come about this way. Many ISO interviewees, for example, would readily tell us “My number is x,” that is, he or she was the xth person hired by the ISO since its inception.

A paraphrased adage is appropriate here: Some operators are born great, most achieve their many skills, and all have the enormous challenge of excelling thrust upon them. “It takes a certain personality to be on the floor,” said an hour-ahead grid resource coordinator. The ideal operator feeds on stress and the added responsibilities of time-sensitive decisions necessary for ensuring real-time reliability. “You maintain reliability through good employees,” said a senior control room manager at one distribution utility,

If they didn't step up to take additional responsibilities, you won't have the reliability you have now. I keep telling my bosses, if we lose one shift supervisor, we're hurting; if we lose two, then we really are hurting. People that are well rested and trained can recognize emergencies before they have occurred. You need someone who understands what they are looking at, not a coaster, but a doer, someone who takes ownership of his job, who enjoys bad days and the craziness.

“Some people like. . .trouble,” said a WSCC security coordinator about control room operators.

Cross-training and shared professional experience have become mandatory, because they give operators experience both with operating at different scales and during different phases of ensuring reliable electricity, e.g., skilled outage coordinators have experience as control room dispatchers (see Perrow 1994 on the importance of such experience for high reliability). Most of those we interviewed came up through the ranks of the utilities, moving from field positions into plant operations and planning and then into headquarters and a variety of upward positions in grid operations, marketing, planning and scheduling. “I was a lineman,” said a PG&E shift

supervisor in its transmission operations center to us, “I think it is extremely beneficial” for being a supervisor. Many in the ISO control room have also come from the same utilities and have similar backgrounds in field operations and head office management. In the course of these career paths, most ISO operators have seen good and bad days, weeks, seasons and times across different levels, departments and tasks over the course of their professional lives.

Such transfers of experience and skills can benefit the receiver more than the giver. “We have lost in the recruitment area. We lost all these people to ISO, when the ISO went up,” a senior manager of a utility’s control room informed us,

All that need for qualified people has created a shortage here. I had two major areas for recruitment: other transmission centers and our power plants. Now that the plants have been divested, now that the generation staff no longer work for the utilities, it took half of the places I usually went for my recruitment. The problems are also worse since the bankruptcy and all the uncertainty, because even recruitment from within transmission is difficult now.

It is not clear, however, to what extent, if at all, ISO reliability has been purchased at the expense of the utilities' reliability. In a sense, restructuring can be seen *as if* it were the result of the three older integrated utilities (PG&E, SCE, and SDGE) having had to consolidate and relocate into satellite market and transmission controls rooms around a new central control room at Folsom—all done by relocating existing staff in addition to new hires across these reorganized control rooms and control areas. If this is true, then it would help explain why service reliability did not suffer as one might have first expected with restructuring, as deregulation was basically the HROs morphing into an HRN, staffed very much by their own people, many of whom stayed in their exact same positions. The individual facts—since restructuring, the ISO has increased its staff from 200 plus to over 500, while staff at some private generation plants and headquarters have decreased during the same period of time—may be less disturbing when considered together within the wider context of an HRN, though this remains an open empirical question.

Whatever the case, clearly the shared background between operators is instrumental in real-time operations and emergencies at the HRN level. “A lot of those in the ISO worked here [in PG&E grid operations] before, we’ve worked together and we know each other,” reported a senior engineer in PG&E grid operations, “June 14 [the date of PG&E’s first rolling blackouts in San Francisco], we did have close communications [with the ISO], we had heads-up like you wouldn’t believe.” A similar sentiment was expressed by his counterpart in ISO grid operations about the level of cooperation on that date.

When asked what makes the ISO control room work, one scheduling coordinator answered: “the fact that everyone who is allowed in the control room is sort of cross-trained in the way they minimally know the functions of the desks.” A generation dispatcher spent nine months on the beep desk and said that had made his job as gen dispatcher easier. The official, who was a member of CERS’s almost ad-hoc created power purchasing team in the ISO control room, had earlier been a control room operator there. A shift manager who had been on a ROPES team course with his crew said, “I’ve got two dispatchers [in my crew], I was one too, so I’m thinking with them, like a third pair of eyes. . .” In brief, many of the operators on the ISO control room

floor have worked on and trained in multiple desks there. Similar cross-training was reported at PG&E's Transmission Operations Center (more below).

Cross-training is important because balancing load and generation has many parts: prescheduling and outage coordination, day-ahead market (congestion management), hour-ahead market, and the transmission desks for inertie generation, beeper, and generation dispatch, among others. Each activity is best undertaken by people who have cognitive maps, on-the-job experience, and real-time institutional memory about the probable impacts of their activities on reliability requirements. More formally, they share the same experiential, institutional and representational knowledge bases. Preschedulers know what the problems are of generators who change their planned outages at the last minute or who do not meet their outage coordination schedules; day-ahead operators know who typically present reliable schedules; inertie schedulers know who they can trust at the other end line to meet generation contingencies. In each activity, the representational, experiential and institutional knowledge represents a composite understanding of what the impacts are on others in the HRN. Pattern recognition in each activity becomes linked through cross-training and experience and the composite amalgam—the shared cognitive map—is precisely what is fundamental to “keeping the bubble,” so important for “just-in-case” performance under real-time reliability conditions

The bubble allows the operator to understand how the control room and the elements in the HRN to which it is in direct communication are performing. This composite knowledge gained through, e.g., cross-training, means that important—but not all!—activities and patterned responses in the HRN are immediately visible or acceptable to the participants, including probable outcomes of current decisions for balancing load and generation “just-in-time.” The bubble serves to make visible operator behavior and their consequences. In so doing, operators become accountable for what they know and see.

One consequence of participants having shared cognitive maps is that this bubble compensates for some information asymmetries in the HRN introduced through restructuring. A common complaint expressed by our ISO interviewees is that the ISO no longer has access to marginal cost information pertaining to generator operation. Yet (1) the knowledge of preschedulers that these generators do meet their outage plans, or have good reason if they cannot, (2) the knowledge of the day-ahead grid resource coordinator that the traders involved do nonetheless bring in balanced schedules and (3) the knowledge of the hour-ahead resource coordinator that certain schedules typically remain dependable, is really knowledge about who is trustworthy—and who is not. “I’ll take a SC that does not give me a hard time and help him out. If he’s giving me a hard time I won’t do that. There are no rules. It’s left up to us here,” says one ISO control room operator. The bubble, if you will, tapers off where ISO operators do not trust other HRN participants or find them unreliable in important respects. Much of the requisite knowledge is necessarily informal and not codified because the surest road to collapse in grid and service reliability is “following the rules” in situations of high turbulence and unpredictability that the California HRN faces because of electricity restructuring and crisis.

Nor is the ISO the only party in the HRN, which is trying to keep its own control room bubble, based on real-time knowledge, accountability and trust. Control room operators in the utilities also place a premium on on-point feedback that is real time and as such more trustworthy, at

least when the objective is “keeping their bubble.” A senior engineer in PG&E grid operations reported in some detail that

we are insisting that for new generation that we have to have real-time information, so that we can monitor generation and transmission. . . We want to know, is it a good day out there? Is that plant coming offline tomorrow, I think I need that information and have a right to know that. . . We don't need/want too much detail, we only want to know enough to do triage here. We want a heads up. That is important for reliability, it is critical information.

We don't want market data, though if the market knows that a 750MW plant is in trouble, that is market sensitive data. So there is info you want but can't have. . . . That relates to trust. Would you like to have the ISO picture? Yes, though more limited. Besides, a lot of that info is coming from us and SCE anyway. There is subset of information that we both need to see. Like the intertie schedules, we need to see and we now do see. But they weren't sending us this in the past. We want to be part of the big picture.

Who the ISO control operators can trust elsewhere in the HRN has become even more important now as a result of restructuring. When asked if trustworthiness had declined as a result of the electricity crisis, a senior ISO engineer answered,

Yes. You see, the rest of NERC is still into the good old boy network thinking. NERC still trusts that stuff to work, so that you do what they say. It's more difficult now. Generator always did what they [the utilities] told them to do. Now they say I got an approved schedule to deliver this amount of power and you [end up with an] overgeneration problem, but that is not my problem. The generators don't have reliability as their objective. Also, NERC hasn't experienced what we have in California, it's beyond their comprehension.

Against such a backdrop, accountability and trust become ways to manage dependencies between participants and activities in the HRN. More formally, accountability and trust have the dual benefit of increasing options variety and/or reducing system volatility that the ISO faces in the HRN due to strategic behavior on the part of others in the network. We cannot, for example, imagine "just-for-now" performance of ISO operators calling up generation operators or utilities engineers and asking them for their cooperation without this trust and accountability being in place. Compare in this regard how a senior generation executive contrasts his unit's official and unofficial communications with the ISO,

. . . the plant control operators do not directly deal with members of the ISO--we use our traders as the link. When you have more than one person speaking for the company, it confuses the issue. The traders are communicating to the ISO the best they can in their own language and the plant personnel relate in another [language].

The Importance of Shared Communications

Feedback, trust, accountability, their underlying reciprocity and overarching bubble are most visibly evidenced through one medium—access to shared communication. The control room

operators we observed and interviewed spend a great deal of their time on the phone. Indeed, the best empirical proof that a “high reliability network” actually exists in California are these phone calls, especially the calls made by control room operators on their direct phone lines and speed dial monitors—which are the quintessential examples of “always on, always available” communications.

Start with the ISO. One gen dispatcher described the ISO touch monitor speed dial system at his desk. He read off to us the outside scheduling coordinators and related units he can call including the: Alhambra dispatcher, Anaheim (a muni), APX, Calpine, CDWR (control room for State Water Project), CERS, Duke Energy (trading desk, the real-time people), Dynegy, Edison merchant, Enron, Mieko, New Energy, Reliant, Pasadena (a muni), PG&E Portfolio and PG&E UES (in actuality both PG&E Portfolio and UES had been consolidated into one unit EPOS), PX (which was no longer in operation at the time of interview), Edison Portfolio, Southern (now Mirant), San Diego Portfolio, SEMPRA (San Diego), WAPA COR (City of Redding), WAPA generation (outage), Williams, and Sierra Pacific.

Now focus on one of the units just listed, PG&E’s EPOS. An operator in that control room told us, “We have speed dial to the ISO gen dispatcher at Folsom, and also to Alhambra for the RMRs...Who do you call most often? Fresno, Rock Creek, Caribou, and others [plants] to a degree. We do have a direct line to the TOC [PG&E Transmission Operations Center], but these calls would probably not be considered routine, they’d have to be related to transmission and our transmission requirements.”

Some of the plants PG&E used to contact directly are now owned by private energy suppliers. In our interview with plant operators in the one of these now privately-owned plants, an operator showed us his direct lines, particularly one to real-time trading floor and traders and the other control rooms at the supplier's other units. The network links come full circle in the fact that, while the ISO gen dispatcher does not have a direct line to the plant we visited, he does to supplier's SC on its trading floor just as does the operator in the plant's control room. In addition, we suspect but did not confirm that the ISO gen dispatcher might still phone the plant informally during peak days.

The HRN extends further, when you move to other ISO desks to their direct speed dial pages. An ISO beeper showed us his direct dial monitor. At the time of the interview (October 2001), most of his calls were to CERS, which had several direct lines. Otherwise, it could be anyone calling him on their direct lines, e.g., Duke, Enron, and Dynegy, which were used frequently. As for PG&E’s TOC, its operators have direct speed dials to, among others, the ISO control room, including the gen dispatcher and beeper, a page of numbers for the munis, another page for support services they can speed dial, and numbers the independent power producers (IPP’s).

Phone calls on these direct lines are the life-blood of the HRN. A shift supervisor in PG&E’s Transmission Operations Center reported at length on the central importance of communication to what we have been calling “keeping the bubble,”

For what we do, everything goes through ISO, and the better the communications with them, the better for you [here]. If that person [in the ISO] has a bad day, they’re

frustrated, then you're frustrated. . . Originally they didn't understand our language, communication was the hard part, and still is in some ways. They're saying it one way and you're hearing it another way, we've worked very hard to iron these things out.

A hundred percent of communications is phone, very little email . . . For the most part, we contact the ISO in all trouble, and with all information on our workload in particular for critical work (less so for non-critical work). . . What you want to do is work with the communication's piece. Everybody needs to be aware and in tune with the outage going on. Things can change and make what is non-critical now critical. . . We are the single point of communication for the ISO, we are their notification process, we keep ISO informed of our part in the system. The people we are in communication with are the ISO transmission people and/or shift manager there. We also can contact the generation desk or Alhambra (which monitors the RMR). The scheduler notifies and in touch with munis and is the single point of contact. .

[Because of restructuring] we have no communication with the power plants, we are not in generating anymore, we don't talk with Moss Landing. Even Hunters Point which is ours we don't talk to. Our people out there [PG&E field operators] contact power plants for voltage purposes. The only place we have direct communication is with Diablo Canyon, the supervisor in charge here communicates with Diablo Canyon.

As communication is so central to so many of the operations in the HRN, it is understandable why the lack of communication is often cast in terms of poor feedback, distrust, and the lack of accountability and reciprocity. We were told on several occasions of run-ins between operators in different control rooms, where, e.g., one operator "would ask for info, to which [name of operator in different control room] would say, I can't tell you that info, it's confidential."

Consequently and it cannot be said too often, there are very real limits on these shared structures of trust, accountability, backgrounds and experiences within the HRN created by restructuring. This report enumerates many instances where the guiding motives of private generators, ISO managers and regulators differ and work against ensuring service and grid reliability through always-on management. Two more differences deserve mention here. First, there are other important differences between parties within the HRN. Two interviewees in the HRN, whose different positions shall go nameless, were candid about HRN relations from where they sat in the network,

Let's face it, the ISO is also a quasi-regulator and your relationship with regulators is very different. Between CPUC and us, there is no trust: It's hard to respect [a senior CPUC official] who has an epiphany about electricity in Disneyland. So on one hand you have the risk averse mentality [of the utilities] and on the other hand you have some of the cowboy mentality [of generators]. So all these things are playing in. Something has obviously got to change now. You need some degree of trust with regulators. With FERC we're ok. . . With the ISO, it is better, but there is tension between us and the ISO higher levels. At operations level, though, is where we get along well.

With the ISO, it's your guilty until proven innocent: They pass the charges onto you and you have to show that they're wrong. When it comes to metering, we'll really be screwed because what's scheduled and what's metered will never match, and look who benefits from the deviation charges—that's how the ISO is making its money.

Not only are there turf issues separating parties within the HRN, but staff in and around their respective control rooms also exhibit contrasting orientations to major activities. The differences between engineers and operators in power utilities is well documented (von Meier, 1999) and we were told numerous times about differences in the ISO control room between operators and crews.

Indeed, the differences between operators, teams and crews within each control room is a very good demonstration, we believe, of the important role of equifinality in ensuring reliability, i.e., not only are there many ways to achieve reliability in real time, but it is vital that there be as many ways as possible, especially during conditions warranting "just-in-time" performance. A senior utilities control room official described this phenomenon and its relationship to trust as follows,

It is really important that my transmission dispatcher and shift supervisor are working closely. I have to ensure that shift supervisor knows what is going on, so the transmission dispatcher has to tell the supervisor everything he is doing until the supervisor builds up the confidence that the dispatcher knows what he's doing. You build up trust. My policy has been to keep a crew together for a minimum of a year—though in the problems we have now I've kept teams together longer—but my idea is to roll shifts every year. Each shift supervisor has his own personality, requirements, training methods, and expertise, so if you rotate learners through these different shift supervisors, then you get a better education for them.

Such views were expressed in the ISO by a senior engineer there who told us the idea there was to try to keep the shift teams together. He went on to mention the special relationship that was needed between the gen dispatcher and the beeper, adding "each team had its own personality, and when someone didn't work, the guy would leave and go to another team."

Conclusion

The casual reader may be tempted to conclude from the preceding that the most pressing design issue facing the HRN is to rationalize and formalize its processes, especially in the area of communications. For some, the ad hoc arrangements, informal communications, personal connections, and patchwork of give and take cry out for formalization, correction and improvement. Operator discretion should be curtailed, control room behavior standardized, procedures followed, phone calls made routine, better protocols put in to place, more automated programming. . .and so on for similar calls to action.

To draw those conclusions would be to draw the exact opposite lesson this report finds in the California electricity restructuring and crisis.

There are many ways to improve the HRN, if redesigners take grid and service reliability seriously, and the major ones are sketched in Chapter 10 recommendations. But the catch is the proviso, "taking reliability seriously." This means is redesigners must ask of each "improvement," How does it decrease system volatility, increase network options to respond to the volatility, enhance the cross-performance adaptability of the focal organization to shifting

performance conditions outside its control, and/or interfere with the pressing need to maneuver out of “just –for–now” performance as and it when it arises? Clearly, if you ask control room operators, what would formalization, codification, curtailment, routinization and rationalization of informal channels of communication do to the flexibility needed to respond to unpredictable or uncontrollable events, they would speak with one voice—grid and service reliability would be harmed, irreversibly. We agree.

The lesson we draw in this chapter and from our research is that what others see as “messy informal communication, unofficial channels, discretion, lack of formal rules and protocols” reflect, on closer inspection, core features we associate with a high reliability network, with the accent on “high reliability.” Such features are close to those that characterize the earlier HROs and include the organization’s constant search for improvement, high pressures, incentives and shared expectations for reliability, and flexible authority patterns. In the preceding chapter, we saw how the latter feature was especially important for the wraparound and control room’s coping with emergency under real-time constraints. In this chapter, flexible authority patterns—the ability of the network parties to go lateral collegially or informally as well as vertical officially and formally across the HRN—have also been found to be absolutely crucial to maintenance of reliability under the same conditions. While the HRN is not an HRO, the legal and political mandates given to the HRN and its parties have never wavered from those of ensuring that service reliability is always-on or only-off when it threatens grid reliability.

In short, society cannot have high reliability critical infrastructures without embedded matrix structures, not only within HRN control rooms and their wraparounds but across them as well. Take away the informal, unofficial and discretionary, and you kill the reliability. Indeed, informal communication and authority structures are so important that they reappear in every chapter that follows, e.g., witness their central role in the operations of markets and market-like behavior under restructuring (Chapter 7).

The preceding chapter ended with a challenge to the ISO that holds as well for the HRN as a whole: Designers—be they politicians, economists or engineers—who are intent on decontextualizing what is relational communication and distorting informal knowledge into representational facts, figures and procedures are the major threat to electricity reliability in this state.

Some may argue that these concerns are beside the point. Over time, with new generation, better information, more sophisticated technology and market redesign, things will settle down. Stability will come. Volatility will decrease, options increase. We think not. The restructured California electricity system is one that will continue to be permanently incomplete and thoroughly unfinished, even when future improvements end up enhancing reliability. This in fact is the most important “always-on” feature of the California electricity system. The system we have created through restructuring will always be a source of periodic increases in volatility and decreases in option variety. We now explain why.

5

PERSISTENTLY INCOMPLETE DESIGN

The Missing Pieces

“So we deregulated and everybody was happy for three years, and then the missing pieces made the thing blow up..... What went wrong in the California deregulation is that different parts of the responsibilities were not located anywhere, such as the responsibility to maintain adequacy and safety.” This senior state government official points to a conclusion that cannot be escaped in hindsight: Restructuring left us with an under-designed system. “All of this happened because the network left a bunch of things unassigned,” he pressed. California is not the only system to face this, argues one of our interviewees familiar with out-of-state utility deregulation: “There is a lot that needs to be smoothed out. Now it’s the wild-wild west. We have got a totally new regime and we don’t know where the holes are yet.”

During the 2000-2001 crisis, many of the holes in the design of the California electricity system became glaringly obvious. By mid-2001, every one was stumbling into or around them. Some ratepayers and consumer groups in the RTE were saying it would be a good thing if PG&E went bankrupt and passed from the scene. Relations in the RRN coalesced into one big vacuum: an ostensible regulator seemed moribund, another in search of a mission, while others were unresponsive. Relations in the HRN looked to be an even worse combat area: the unremitting strategic gaming; the PX here one day, gone the next; PG&E declares bankruptcy and then CERS shows up in the ISO control room. Markets became turbo-markets, prices hyper-prices, and the area control error (ACE) at times hardly controlled at all. Operators went to work in the morning to find they were 10000MWs short of meeting that day’s forecasted load. One day we sat in a meeting where it was seriously touted whether the ISO would become part of CERS, or the new California Power Authority (CPA), or whatever.

Nor have the currents in the restructuring pond ceased to move outwards since mid-2001, as we watch deregulation and Enron *et al* swell into a national security threat. "Several recent trends in the energy industries have increased the vulnerability of their infrastructures and made serious loss of service from terrorist attack more likely," states the recent National Research Council report, *Making the Nation Safer* (2002, 6-2). Surprise! One of those trends is deregulation. "This economic and competitive setting has led to reduced investment in system capacity and technology development...[putting the grid under] increased stress even without the threat of terrorism." The report concludes, "Deregulation has encouraged efficiency, invested-cost utilization, and return on investment rather than redundancy, reliability and security [in the national grid]" (Ibid, 6-2, 6-3, 6-4).

All of this and more has given rise to the conventional wisdom that the electricity crisis was the result of poorly designed institutions, policies and legislation. For many, the problem looks to be incomplete regulatory oversight, gaps in long-term planning, poor market design, misaligned business models and more. Our evidence uncovers two major mechanisms that led to and will continue to renew the incompleteness of the design of the California networked electricity sector. They challenge the notion that we are in transition to a more or less finished design that will stabilize performance conditions for a more predictable and reliable electricity system.

Incompleteness in HRN, RRN, and RTE

We define an incomplete design as one that is missing crucial elements for it to function in the way it was originally intended. Of course, just as there is no perfect information, no perfectly complete design ever exists. Trite but true, there are always surprises and changes along the way. That said, our interviewees agree that crucial design elements were missing from the outset. Restructuring was a train wreck that did not need to wait very long before happening.

The crisis brought out many gaps in the HRN design. It is not our intention to identify them exhaustively, but to illustrate where and in what form the design proved to be importantly incomplete. First on most people's list are the much-maligned market designs that produced all manner of unintended results. Three experts (Chandley *et al*, 2000, pp. 2-3) summarize what they see as the "fatal flaw:"

California built its market design on a flawed premise. It is a commonplace that electric systems are both complicated and highly interdependent. Over short horizons of a day or less, generating facilities must work through the transmission network to provide the multiple products of energy, reserves and ancillary services. The same generating facilities must provide all of these products, in the right amounts, and with very limited tolerances. The simple physical reality dictates that these services must, in the end, be coordinated by a system operator. There is no other choice available with our current technology, and every electric system has such a system operator. The flawed premise of the California market design was that this inescapable reality could be ignored or minimized in an effort to honor a faith in the ability of markets to solve the problems of coordination. This was an unprecedented experiment in markets that did not work in theory. We now know that it did not work in practice either.

This general observation can be easily documented by anyone familiar with HRN operations: congestion management lacking adjustment bids and being gamed through fictitious load; unrealistic balanced schedules submitted to the day-ahead desk; underscheduling, overgeneration, and lack of sanctions for not following dispatch instructions (or rather too many incentives not to follow dispatch instructions); generators selling the same output twice in different markets or units not showing up at all; "kilowatt laundering" out of state to escape price caps; and the list goes on. These examples are not so much transgressions of the design—although that is undoubtedly going on also—but rather the exploitation of the structural holes and gaps in design.

Some problems were foreseen, others were more surprising, but none were addressed adequately beforehand. Nor could they have been thoroughly addressed beforehand, in the view of an

increasingly number of observers. Having attended a conference of the Electrical Power Research Institute, a senior ISO engineer concluded, “a lot of blame goes back to gaming the particular design we had. What I learned from the EPRI conference is that all market designs would have led to some inherent flaws in the electricity markets. . . It always appears as a flaw in a particular design, when in reality it is in every design.” David Freeman, chair of the California Power Authority, has gone so far to say that deregulation itself is “inherently gameable,” according to a report in the *Financial Times* (May 16 2002). As an ISO hour-ahead grid resource coordinator explained, “gaming on part of SCs is annoying, but I can’t do anything about because it’s not illegal and they are pretty blatant about it.”

Another missing element in the design of the HRN was outage coordination. The original design required only RMR units to coordinate outages with the ISO. That left the major part of generation unaddressed, as a senior ISO outage coordination official expands,

All [non-RMR] generation, which is about 80%, just had to give us an annual plan. Technically, after they had submitted that plan, that was the last time they had to call us, unless they changed the plan. Most of them do call, however, because that is what traditionally was done. We don’t have authority to deny them outages, but we can ask them to move the outage informally. That [lack of authority] hadn’t been a significant problem until December last year, when 3000MW was out of use in nuclear, plus others were out of maintenance to get up for the summer... [We] had emergencies in December, which was unprecedented. The network idea does work with exchanging information, but not if the generators can’t or won’t move because of other constraints [like] their business. We were postponing outages and we think we had some control over it through political pressure. We were making it very visible that the shortages were because of generators being offline, and during that time FERC said that ISO’s control of outage coordination had to increase.

Transmission maintenance also “fell through the cracks.” A senior CPUC staff member argued that “the incentive structure was to walk away from maintenance. The ISO did have an interest in it, but had neither the resources, nor the authority, nor the institutional structure for inspection. We [CPUC] were supposed to do that... But we never had the resources for spot checking this in the field. This was not put in the contracts. We don’t do maintenance; we were just trying to make sure if the utility was doing that work.”

A blatant, yet scarcely acknowledged, missing design element are the environmental safeguards in operating the electricity system. Whatever provisions were in place before the crisis, they were put on hold when the crisis accelerated, as witnessed by increasing exemptions to emission restrictions and the push for more natural gas generation displacing an enhanced commitment to renewable energy. During the crisis, we found trace elements of environmental concern only in a few daily operations within the HRN. An ISO engineering manager explained how their proxy bid system was hampered by the fact that “federal air quality and FERC guidelines don’t mesh and in some cases are in conflict with each other.” An ISO operations engineer working on RMR contracts told us how “One day I got call from an owner saying, ‘I can’t produce because of this or that restriction.’ It started with Delta [smelt] dispatch. During spawning, larvae are below a certain size for Delta smelt and striped bass, which means that the Pittsburg unit in the Bay Area can’t go above 86 degrees according to a protocol and then only up according to prescribed

steps. So I had to translate all of these protocols into ISO procedures.” A senior CEC staff member told us that in a much broader sense such concerns were missing from the design,

The ISO in California is making all the market decisions and transmission decisions based on economics. There are no provisions for environmental or health and safety policies or objectives. There is no broader public policy framework... Traditional market rules may not be applicable to that. They do not appear applicable. And that is disturbing.

Many missing pieces in the HRN are mirrored in the RRN. Regulators were created that initially had little to do (the Electricity Oversight Board), while some of what needed to be done was left unregulated. Oversight responsibilities, for example, were not clearly assigned or in ways that made them very difficult to implement—as in the case of transmission maintenance. Ironically, the restructuring created more regulatory incompleteness rather than less by going from two entities to six counting. The CEC staff member explained:

Before restructuring there were two agencies that dealt with energy in California. That was the PUC, that dealt with rates, and the CEC that dealt with permitting and forecasting. In restructuring they ended up creating five entities with the EOB, ISO, PX. The other thing they did was that they gave a major responsibility to the FERC, for the wholesale market. And that wholesale picture became much more important than it was before. The Feds have a different picture than California. They have a different perspective, different policies and different objectives which may not be the same as California has. That was a major change. The other thing that happened was that by creating all those new agencies, the roles and relationships between them were not very clear. There wasn't a very clear demarcation of who was responsible for what. The legislature last week passed another bill, the Power Authority bill and that created another entity. Now we have six. We haven't really clarified those roles and responsibilities.

High reliability theory has identified close oversight by outside bodies as a crucial feature of high reliability (Appendix B) Contrast this, then, to the picture painted by a senior legislative aide illustrating the shifting and unclear oversight for reliability:

The CPUC is in a kind of struggle with providing reliability. They still regulate the utilities, but they only regulate the distribution service and part of the transmission service. But private parties own a lot of generation and CPUC has no economic control of these generators. Over the last months, we have had unplanned outages, due to maintenance problems, accidents, and a lot of people suggest that private companies were withholding power to raise prices up. CPUC really doesn't have clear mandate to follow that and sanction it. Their ability to ensure reliability has diminished as well as the utilities' ability to maintain reliability as generation has been privatized.

Even with so many entities—or rather because of so many entities—certain responsibilities were “walked away from.” “Nobody took responsibility for prices,” says a senior state government official. FERC remained reluctant to intervene on the wholesale side, even after repeated requests from California to do so. CPUC had a similar position on the retail side. From their respective mandates, these positions may be understandable, though their implications continue to be far-reaching. CPUC's primary concern is for the utilities' customers, which meant that it could not simply relinquish the mandate to assess whether forward electricity contracts were fair

and reasonable. That position contributed to the unprecedented shifts in electricity transactions within the ISO markets, most notably to its real-time imbalance market, as the state official explained,

The ISO has had a reliance on the spot market, not on forward contracts. CPUC will argue that, sure you can do forward contracts, you're allowed to do it. But it also says, you can forward contract, but we (the CPUC), with 20/20 hindsight, keep our right to assess whether these costs were fair and reasonable five years from now. So the market participants say, I'm not going to do that. They went into the ISO markets, where they faced no such risks of assessments, they have no risk in that market. In the ISO markets, they can buy and pass on the costs without questions, and they did that big time.

Gaps also occurred regarding provisions for long-term mandates. Prior to restructuring, the state linked siting and energy development through integrated resource planning, following from the legislation creating the CEC. Restructuring changed all that. According to the legislative aide, "The bill we did a couple of years ago basically severed the connection between siting and planning functions. We are rethinking this now, if we go back to integrated planning."

Last but not least, we found design to be incomplete in the way the HRN and RRN relate to the RTE. A crucial missing element is discussed in the next chapter: For the task environment, the reliability of electricity is non-fungible—that is, it is socially and politically impermissible to trade it off at the margins with other values, especially in real time. This has been illustrated all too dramatically by reactions of the public, the California economy and the political leaders who felt compelled to step in after the windfall transfer of California wealth to Texas. As one of those involved said, "that was one thing we failed to think about." It was not the only thing.

"We Are In a Transition Right Now..."

The most frequent statement in our interviews must be variants on "We're in transition now..." The statement marks an important policy narrative in the field: Yes, the design was flawed, but we're trying to fix it, fill in the gaps, finish the job. Once that's done, volatility will decrease and the system will work like it was supposed to—reliably. "Right now, we are in this transition period," said an electricity restructuring expert. "The people that are running the system are still trying to run it in the old-fashioned way. That doesn't work anymore. They have to be more responsive to prices now. I try to explain this to engineers." An ISO grid resource coordinator expands the point: "A lot of the problems, these growing pains, ... will go away once more generation comes on line." A senior generation executive of a state energy supplier concurs: "In the long term it might be more reliable [as more generation comes on line and stabilizes the situation], but we are definitely in transition now."

Certainly the system is in transition, but the question is: What comes next? The problem is that the policy narrative is systematically misleading in its implication that once the design is fixed, things will be "normal." That hoped-for event coincides with another narrative at large: The 2000-2001 crisis was the perfect storm, the unlikely, unfortunate and highly contingent confluence of low snowfall in the Pacific Northwest reducing electricity imports, lack of new generation in California, fast-growing energy demand, high natural gas prices, and poor market

design. What are the odds of us facing the same coincidence of factors twice? The (tautological) answer: It is highly unlikely. The hope to be embraced is thus the same. If we can just get through this, things will revert to normal.

What is needed, in the words of one of the ISO market managers is “stabilization, meaning that once a market design is implemented, the rules and associated procedures for participating in the market could remain unchanged for a while, so that everyone has time to learn the rule.” Our interviews and research lead us to conclude that this will not happen any time soon, because structural factors put into play by restructuring are ensuring persisting incompleteness in the design of the electricity system. There is no stable end state, as volatility has now become a recurring feature of the California electricity system. In the words of one expert, “The big mystery for me is the long run. There will be a tendency to have long-run forward contracts, but I do not know what the balance will be between long term forward and real time contracts. Maybe we will always have a balance problem.”

From a reliability perspective, we all must ask how can the system deal with persisting incompleteness. It also means not postponing provisions to ensure reliability to “after the transition” on the hope that we can muddle our way through until then. It further means we simply cannot assume there ever will be a design that keeps volatility, options and adaptability in balance through “normal” means. To see why, we turn to the structural factors underlying the design incompleteness.

Persisting Incompleteness

Our evidence uncovers two forces driving the incompleteness of the design of the California electricity system: the complexity and the dynamics of any large-scale high reliability network. Whenever such a network is established, whether under deregulation, re-regulation or whatever restructuring effort, it never fits all together. Gaps, mismatches and unintended consequences inevitably emerge, some as insurmountable obstacles, others as unexpected opportunities to be exploited. Not only is complexity an emerging property, but the sporadic attempts to rectify the unfinished designs are also ultimately overtaken by dynamically changing events.

Incompleteness persists because redesign efforts and adaptations are always the provisional patchwork of divergent, at times polarized, interests. Some interests may be locked in stalemate over proposals that try to correct for design incompleteness. Or others may generate a compromise that leaves all manner of design and management issues systematically and intentionally unaddressed, giving rise to new problems of design in the process. The latter is basically the mechanism that produced that deregulation compromise in the first place which the headlines and public contempt continue to return.

Take a closer look at the forces. Few people would disagree that the policy and institutional design is complex. One key challenge of complexity for the HRN is the match between technology and market. Both are complex enough on their own, with their own design logic, quality standards, management requirements, expert professions, and field patois. Yet if markets are to coordinate and guide grid operations, then the markets have to reflect the physical properties of the grid. The laws of physics trump the laws of economics, unless of course you believe a contemporary version of the medieval theory of double faith, which held that Aristotle

was in heaven right up there with Jesus. In a perfectly competitive electricity market, it is possible to imagine transactions occurring in real time without the need for operating reserves; in a world with a real grid and electrons, it is impossible to imagine not having operating reserves, if reliability is the mandate. One of the ISO's lead officials in market operations summed up: "You have to follow a physical model for operating the grid. However, to facilitate markets, we allow commercial models in the forward market. We're trying to find that mesh between commercial and physical, which is one of the biggest concern of ISO... Somehow the two models have to meet, the question is where do they meet: in real time or in the day ahead. This is the tradeoff between reliability and facilitating markets." That has been the unremitting challenge faced by ISO market engineers and other experts involved—"I don't have a lot of good days," said the market operations official.

The multi-faceted design that restructuring gave rise to is difficult enough to manage, let alone redesign. "I look at the dispatcher log and their list of what went wrong [the day before in the ISO control room] and think, this is not acceptable," reports a senior ISO engineer, "We are training dispatchers for a year and still we don't have the right training to get people out there [on the floor] with better skills." According to a manager of real-time scheduling in the ISO, "With deregulation, there is a constant stream of people—new, bright, but unfamiliar with standards and scheduling."

Scheduling, in particular, followed the rules of the market, but created new problems in grid operations. The market operations official: "There are no checks [for congestion] within the control area beforehand. The point is that the SC can give me any schedule on supply and demand. . . That is why we have failed." In a similar vein, many provisions to connect market and technology face tradeoffs between accuracy and transparency. Accuracy is needed to reflect the complex physical realities of the grid. Transparency is needed to have markets that present participants with clear price signals and enable economically rational behavior. More accuracy is generally bought at the expense of transparency and vice versa. The ISO's knowing that a generator has a boiler leak increases accuracy of the ISO's outage coordination at the same time the transparency of such information to other generators may work against the ISO in terms of the real-time prices and strategic behavior it now has to confront from these remaining generators. "If the ISO comes in and does black box type things, it is not transparent to the market participants," the market operations official explain. "So do you want accuracy or transparency? Reliability needs accuracy, but is not as good for transparency."

Obviously important improvements can and have been made. "Nobody envisioned gen dispatchers would have to make the number of calls they actually had to make," according to a WSCC security coordinator. Consequently, as pointed out previously, the Automatic Dispatch System (ADS) was introduced and reduced that workload significantly for the dispatcher. Unfortunately, the same conditions that press for improvements are those that create the opportunities for design errors. Under the high volatility performance conditions of real-time reliability, design misjudgments must be expected in the control rooms and within the networked workarounds. "Here's an example of one misdesign," said an expert.

The ISO has had four markets: spinning, non-spinning, one hour non-spinning, and replacement. ...[F]or each type, a completely different market was created. The result

was that when they were running out of replacement, they would have to go out and buy really high-cost replacement to make up the shortfall. So instead of substituting replacement with, say, spinning reserve, they ended up paying \$10,000 for added replacement rather than substitute high quality spinning reserve at, say, \$30.

In organization theory, one chief manifestation of design complexity is surprise, and the surprises have been many in the restructured system. The numerous gaming strategies within the HRN—some predicted, others creative or simply unexpected—have been repeatedly commented upon. Yet more rules are not necessarily a buffer against surprise. A senior CPUC staff member: “There are a lot of elaborate rules about pricing. Notwithstanding the rules for different markets and price caps, if the generators wanted money, they’d get it.” When asked whether our interviewee had foreseen the current problems that the ISO is undergoing, the electricity expert admitted “we saw a lot of small inefficiency issues. I did ask about retail versus wholesale prices, how do they really know wholesale prices won’t go up, but no one expected the kind of large increases we see now.”

The complexity of the restructured system produced other surprises. A senior engineer with ISO explained how the qualified facilities (QFs) were a source of the unexpected in estimating available generation for use by the ISO,

No one was on the hook to provide generation. ISO said all along that we weren’t getting the generation... In reality we were seeing 38,000 [MW] and plus five or six thousand in imports, so we had 43-44,000 [MW] load, while the spreadsheets from CEC showed 46-50,000 MW of nameplate capacity—literally, photographs with brass plate of generations... Actual data in January showed, contrary to the media reports that generators were offline, that the generators were online and were showing up, and they were...working hard. It was the QFs that had dropped out.

Some events are not so surprising to some insiders. They know the reality behind the problematic numbers. That does not mean, however, that the problems are actually addressed, as one shift manager explains: “Non-firm load is the power you are receiving in your control area and need to cover yourself with reserves. However, a lot of schedules that we assume are firm, are actually not firm. It’s kind of a ‘don’t ask, don’t tell’ policy. The WSCC knows about this, and can’t tell itself what load is firm or not. Nobody wants to know.” According to a senior ISO engineer, “we are starting to count interruptible load as OR [operating reserves], even though that technically is not right. Overall, the enforceability system doesn’t work as good as it should have and could have.”

Surprises stem not just from complexity, but also from the continually changing conditions under which the system operates. Dynamics are the second important driver of incompleteness. The California system “has a dynamic instability problem as well as newer voltage instability problems related to reactive power and voltage issues,” an engineering expert informed us. “This kind of problem is new because of market operations.” That said, the system was inherently dynamic even before restructuring. Working on “Adaptive Infrastructures,” EPRI’s Massoud Amin (2001) identified an array of forces working on the system in addition to restructuring, creating an “increasingly stressed network.” These forces are putting new demands on the design, creating new gaps and overtaking attempts to put into place missing pieces. To name a

few: urbanization, challenges of the “digital society” for higher reliability, right-of-way restriction for transmission and distribution expansion, new environmental constraints, increasing public resistance to power systems as neighbors. A Department of Energy (DoE) electricity expert concurs that the system has been changing: “The increase in the utilization of the system for economic trade started long before deregulation. But deregulation has dramatically increased the volume and magnitude of transactions.” A senior grid operations engineer at PG&E described how the grid, by becoming more dynamic and complex, was “now operating closer to the margins.”

Unfortunately, redesign efforts themselves are a major source of in-built incompleteness. In one sense, tensions and flaws within and between the HRN and RRN cry out to be addressed if not resolved. How else are we to achieve and keep reliability without ironing out the tensions between regulators, such as CPUC, FERC and the Governor's office? How else are we to achieve and keep reliability without ironing out the differences between the ISO and generators, between the ISO and the distribution utilities, between the generators and the distribution utilities, and between the HRN and the RRN participants? But these questions of course raise a much more sobering one: Why should we expect the RRN to redesign the HRN for the better, when the regulators haven't been able to rectify and redesign their own RRN?

The conclusion is clear: If the system is intrinsically dynamic and emergently complex, its design must be persistently incomplete in important but unpredictable respects. Yes, efforts to redesign can and must improve HRN performance conditions. What they cannot do is promise stability “after this transition.” In short, we conclude that planning and managing for the reliability in the California electricity system would be on a much sounder footing and much less vulnerable, if the planners, managers and operators were working under the assumption that design incompleteness and induced volatility will persist for the foreseeable future. The question then becomes: What are the implications of persisting incompleteness for grid and service reliability?

Negative and Positive Implications for Reliability

Design incompleteness has negative impacts for service and grid reliability, certainly. The many examples in this chapter and report make that deeply clear. Much less understood is that incompleteness entails under-acknowledged positive implications along with the obvious negative ones. For many of the negative impacts of design incompleteness have assumptions built on sand rather than into rock.

In the first place, it is a highly questionable assumption to believe that every thing would have been alright if we hadn't simply slapped things together so quickly at the start with restructuring, if we simply had had better integration, better coordination, better politics, better markets, better people. . .

Whatever the psychological appeal of this utopian prospect, it is not supported by organizational research or the literature as we know it. The generalization that more coordination, less compromise, better integration, happier people and clearer responsibilities are everywhere superior to messy networks of multiple, overlapping, conflicted and fragmented organizational

units is simply not supported by the empirical evidence available. There are advantages and disadvantages to design incompleteness and several of the more important factors are summarized in Table 5-1.

The left-hand column of Table 5-1 lists the recurrent problems of incompleteness. Notwithstanding the many gaps in reliability coverage, many regulators in the RRN do have assurance of reliable electricity as one of their mandates in some shape or form. This duplication, of course, creates and supports overlap in staff, projects, research, and expertise, all of which indeed utilize more resources. That is, given the complexity of the system being managed or intervened in, the duplication is also much-needed redundancy for the reliability of large scale technical systems (see Landau 1969; Lerner 1986). This and other positive features are listed in the table's right-hand column.

Fragmented mandates increase the dependency on interagency cooperation and ensure that many RRN officials review the same projects or policies—a practice that can be cumbersome but which allows for functional specialization and economies of scale regarding the organization of expertise. Should a single regulatory agency have the unencumbered mandate to review the entire system design, it would have to develop all areas of expertise now provided by the set of specialized programs and staff. Conflict has long been recognized for its functional value in large scale systems. While turf battles are sure to make all redesign efforts difficult undertakings, they also provide the best possible guarantee that different views and interests—as long as they are institutionalized—are taken into account in the design of solutions. Put in these terms, the positive role of incompleteness clearly is to keep polarized issues in the public arena and under scrutiny. It increases the transparency of those tradeoffs, keeping them from being made implicitly inside the “black box” of programs and agencies, whose members may seek to fold or otherwise obfuscate conflicting issues under one mandate.

**Table 5-1
Positive and Negative Implications of Persistently Incomplete Design**

Feature	Negative	Positive
Overlapping mandates	Wasteful duplication	Positive redundancy
Fragmentation	Segmentation of authority, some responsibilities left unassigned	Functional specialization, economies of scale
Conflict	Turf battles, stalemates, compromises	Guaranteed consideration of different interests, debate over competing proposals, keeping tradeoffs in the public arena
Unintegrated priorities and goals	No system-wide priorities & tradeoffs	Institutional protection of vulnerable goals & interests
Accountability	No accountability for overall performance	Accountability directly related to specific tasks
Disjointed information	Lack of coordination in information gathering & assessment	More error correction, less distortion in aggregating or assessing information
System redesign	Governed by compromises, unable to achieve necessary systemic redesign	Incremental, goal-seeking redesign

Having competing goals embedded in different agencies and entities obstructs identifying tradeoffs and setting priorities among them during redesign and implementation, especially those that translate directly into issues of authority and funds of the regulatory and operating agencies concerned. Coupling the goals through one design, however, may not bring striking well-balanced tradeoffs any closer, particularly if some goals are less firmly institutionalized and more vulnerable than others. A prime example of the latter is the goals of long-term reliability of electricity, which is underwritten as a key goal by several regulatory agencies. But in the process of day-to-day tradeoffs among competing priorities, that is, between short-term, or even real time and long-term reliability, the latter routinely suffers. (These issues are explicitly addressed in the next chapter on the non-fungibility of short-term versus the fungibility of long-term reliability in setting the tradeoffs between grid and service reliability.)

The issue of accountability can be argued along the same lines. True enough within the current unfinished design, no one is directly accountable for the reliability of the whole California electricity system. Establishing an accountability structure for overall performance may address some issues, but will undoubtedly increase the distance—loosen the ties—between accountability and specific tasks. The latter have clearer, more tangible performance measures

from which an organization derives more direct feedback and learning, different from the aggregate measures of success for overall accountability. Contrast this with the case of the ISO, reliability's pivot in the HRN. Its performance criteria for reliability reflect not only the ISO's own performance, but also events and performance of others in the HRN over which it has little control. Thus the ISO has ended up paying substantial penalties to the WSCC for not meeting WSCC standards on reserves, a problem that the ISO could scarcely be held accountable for during the crisis. Are things better because of the enforceable performance standards, we asked a senior engineer with the ISO. "Yes and no," he said. "The problem is we, the ISO, are paying big penalties because of lack our operating reserves. Are these penalties helping me to do a better job? No. Is it helping us learn to operate better? No." This one-sided experience with penalties and fines, however, does not stop many of our ISO interviewees from calling for more and better enforcement of fines and penalties on the private traders and generators as a way to improve *their* reliability.

Information gathering within the network setting may also benefit from closer connections between information gatherer and user. Notwithstanding the obvious advantages of coordinating information gathering, such efforts can be prone to error and systematic distortions as well as making it more difficult to act upon the information. Less coordinated programs provide more checks and balances against distortion and bring the information user and the information gatherer closer together. Everyone overhearing or overseeing multiple channels of information and communication in the control room—no matter how "extraneous" some of the information may appear to a would-be information coordinator—is key to networked reliability, as we observed during the electricity crisis.

To summarize: The lack of a comprehensive, integrated redesign to deal with the linked problems of the current system has often been cast as the muddling through of compromises plagued by "politics." If everything is politics, then the systemwide redesigns advocated by some—deregulate more! reregulate now!—are non-starters, if simply because politics would surely warp such comprehensive initiatives as well. However, in light of the positive features of incomplete design, a more compelling case can be made anyway for incremental and emergent change, which is not only a realistic model of design of critical service provision, but according to well-known policy experts the more rational one in a networked setting of divergent interests (Wildavsky 1979; Lindblom 1990).

The positive side of persisting incompleteness is that it fits better with adaptive processes within a network structure. Adaptation does not require every change to be formally rationalized and integrated into the whole, thereby purchasing flexibility at the expense of consistency. Where design-based change puts a premium on logic, incremental adaptation and self-correcting emergence are more amenable to acting upon and in light of experiential, institutional and representational knowledge acquisition that is by definition difficult if not invidious to formalize. Such hard-to-codify knowledge bases are even more important under current conditions of design incompleteness, because these force the ISO and other units in the HRN to work outside analysis, as we see in "just-in-time" performance. When working outside analysis, that is, when operations are by necessity improvisational, the ability to act flexibly on experience and localized knowledge is a reliability-enhancing feature.

Or to put it another way, the positive features on the right-side of Table 5-1 are precisely those we would hope a functioning network of divergent interests would exhibit if given responsibility for an infrastructure as critical as electricity.

Conclusion

The positive and negative face of design incompleteness carries a double message. First, there are positive reasons, i.e., incentives, for “unfinished business” to exist and persist. It is not just the outcome of “politics,” “misdesign,” “compromise,” “lack of expertise,” and so on. Second, reliability is difficult to achieve under conditions of incompleteness and the dynamic volatility it produces, but the latter do not make the former impossible. For the California electricity crisis has made one thing clear: Even under highly volatile conditions induced by a restructuring missing all manner of design elements, the system was remarkably reliable.

Does this mean we should be complacent? Obviously not. It does mean, however, that any improvements to the current design must pass what we have been calling the reliability-matters test of whether they would reduce system volatility, increase network options, or enhance ISO cross-performance adaptability. We argue in the report that keeping the lights on during the crisis was in no small part due to the ability of the ISO to adapt across different performance conditions and adopt different performance modes in balancing load and generation, most notably for achieving real-time reliability in terms of "just-in-time" and "just-for-now performance." A significant part of this adaptability is necessitated by the fact that grid and service reliability are so important that they cannot be given up in real time. We turn now to that topic, the non-fungibility of reliable electricity

6

NON-FUNGIBILITY IN REAL TIME

Not Just a Commodity

California spent over \$20 billion for the world's best demonstration that electricity is not just another commodity like corn syrup. The reason why electricity is different is directly tied to its reliability as an always-on service, which has daunting implications for the institutional design of any large-scale electricity system, such as our state's. Some economists claim otherwise, as one did during a recent conference:

I disagree with anybody who sees [electricity supply] as something more than just a commodity. The reason it is treated differently is political only. What actually happened was that we had the worst combination of weather and other factors and that led to political crisis over electricity, when in fact there is nothing different with electricity than with corn syrup.

What economists sometimes externalize as “political,” i.e. non-economic, others insist to be intrinsic to what defines electricity as an always-on service. “Electricity is unique in two respects,” a senior state government official told us,

First, it is non-substitutable. Yes, it's possible to use gas for some machines, but that is limited. It's not like going into a bakery and wanting French bread, but choosing bagels. In the case of electricity, the only substitutability is the decision to consume or not to consume... Electricity is a right, according to our policy, because of the non-substitutability. It is part of the fabric of our existence. Second, it's a right to have it at a reasonable cost, so that it doesn't destroy people's lifestyles. In California we have not addressed the choices around that.

While we may quibble with terms, the interviewee's point is clear: The properties demanded of electricity—its large-scale provision must be safe, continuous and available—are treated as an unavoidably high reliability mandate. In the words of a state legislator specializing in energy policy, "When we look at how we use [electricity] we see there isn't a very good substitute. It is essential for water purification, hospitals, a whole host of communications, it can't be stored, it has to be delivered in real time, and it is far more subject to market power." Important organizational and institutional implications follow, as has been repeatedly shown in the requirements we impose on high reliability organizations (HROs), such as nuclear power reactors, and on the critical infrastructures in which they operate, in this case the US electricity sector. Research has found that HROs share a number of demanding characteristics in order to perform reliably. These features were introduced earlier in Chapter 2 and are outlined more fully in Appendix B.

This chapter focuses on that exceptionally important HRO characteristic within a networked setting: Reliability is not fungible. Because of the catastrophic consequences of error or failure, HROs do not have the option to always make marginal tradeoffs between increasing their services and the reliability with which those services are provided (Rochlin 1993: 16).

“Reliability demands are so intense, and failures so potentially unforgiving, that...[m]anagers are hardly free to reduce investments and arrive at conclusions about the marginal impacts on reliability” (Schulman 1993b: 34–35). Issues of efficiency and cost-effectiveness are difficult to address, let alone assess.

If emergencies are by definition neither cost-effective nor reliable, is it not better to judge an effort that precludes them first by its reliability than by its cost-effectiveness? Indeed, there is the sense that the ultimate cost-effectiveness would be to preclude catastrophes from ever happening. That is, there is a point at which the HROs are simply unable to swap reliability for other desired attributes, including money. These organizations operate under a logic of precluded events, that is, some events have to be avoided at all costs because their consequences far outweigh the costs of always having to preclude them. Money and the like are not interchangeable with reliability; they cannot substitute for it. *High* reliability is, “simply,” non-fungible for these critical infrastructures. “Energy is such an essential heart of our system,” one state energy planner told us, “It is a thing that is unlike any other commodity.”

The treatment of reliability as a non-fungible feature of electricity has defined the California electricity crisis and most prominently the system’s primary reliability-seeking and -enhancing organization, the ISO. In the words of a senior ISO engineer, “The ultimate goal here is reliability. NERC’s concern is reliability at any cost.” While this overstates the case (later we show that “at any cost” no longer holds over the longer term), it is clear that during the crisis, the ISO was not in a position to trade off the exploding costs of real-time energy to keep the lights on against any material reduction in reliability through sustained shedding of load, i.e., through controlled blackouts. “We don’t discuss price, we don’t do market, we do reliability, which is really outside the market,” was how one shift manager at the ISO phrased it. “Here, interrupting load is like the surgeon losing a patient.” A security coordinator working in the ISO control room told us, “You have to understand that this grid system is dangerous, people are going to die if it isn’t managed right. Sure you can deregulate telephones, but no one died because of that.” In the ISO, as with HROs, the commitment to reliability is warp and woof of the focal organization’s culture. Another ISO shift manager felt

I think ... the only way to maintain reliability is to have people on the floor absolutely committed to doing that. This means people on this floor who are not willing to cut corners or be politically influenced, people who know there is a way to operate. . . . Anytime you compromise that reliability you end up with problems. The people here need to have a buck-stops-here attitude.

So strong is the commitment to service reliability that the few times controlled blackouts actually occurred during the crisis were the result of the only real-time tradeoff service reliability actually faces: namely, that with grid reliability. Controlled blackouts were only acceptable when they prevented even greater risks to reliability, those of uncontrolled blackouts and islanding the grid itself. Service reliability that threatens grid reliability threatens the always-on rationale of that service.

The non-fungibility of service reliability except when pitted against the non-fungibility of grid reliability is most tangible and acute in real time under "just-in-time" and "just-for-now" performance conditions of high system volatility. A senior programmer in the ISO market operations observed, "When there are not enough energy bids in real time, operators are forced to go out of market. At that point, they may or may be making the most economic and transparent decisions, because the number one priority is maintaining grid reliability." An ISO grid resource coordinator at one of the ISO's market desks said "Yes, there is all this chasing the load and holding back resources until the last minute at the highest prices, but there's a limit we don't go below. We'll go through the stages and even blackouts, but we won't compromise the system." "It's we [the generators] who tell [the trading floor], we give them the rules, we can tell them how fast to raise load, we tell them we can't do that. We can veto, which means it is our problem, we are responsible for the operation of the units, regardless of commitments made by the trading organization," insisted a senior generation executive of a state energy supplier. In fact, grid reliability of plant and transmission supercedes service reliability, even when margins are not tight. A senior ISO official responsible for scheduling told us that they had to go into controlled blackouts because of a reliability limit on Path 15, even though there was sufficient generation in southern California to meet the load.

Treating grid and service reliability as essentially non-fungible is not the same as meeting all reliability standards set by the RRN, such as WSCC criteria. The non-fungible part of reliability is defined more narrowly: keeping the lights on. A senior ISO official explained that "[according to] WSCC criteria, if we run into red lines [the bandwidths within which the ACE is suppose to remain], we should be shedding load. But we ask ourselves at what risk if we did meet these criteria and caused lights at intersection to go off... We had our hand on the button. But you want to make sure that you don't launch the missiles unless you absolutely have to. Most people have never had to make a decision like that, but when we shed load, lives are at stake." Under pressure of the crisis, both grid and service reliability were redefined, a topic to which return in Chapter 8.

Defining reliability first and foremost in terms of keeping the lights on has implications for the non-fungibility of reliability. In the older, HRO-like integrated utilities, distinguishing between grid and service reliability was all-but-moot, as they were virtually synonymous. High reliability in HROs resides in protecting against errors, e.g., when having to shut down the reactor, go offline, or undertake a job that requires thinking through every contingency beforehand. In HROs, the option to stop is crucial to reliability. In the HRN, service reliability depends crucially on the ability to keep going. That is, service reliability resides first and foremost in continuous action to maintain or restore the balance of load and generation, where there is no viable shut-down option to maintain that reliability. Only when faced with a possible grid breakdown does reliability shift to shutting down service in the form of shedding load, but even here the goal is to prevent grid-wide islanding. In the HRO, being reliable entails the ability to stop the service and know what will happen when that is done. In the HRN, being reliable requires keeping the service always on and not knowing exactly what is going to happen—until, that is, grid reliability is jeopardized. To summarize, in an HRO such as Diablo Canyon nuclear reactor, the option to stop is non-fungible. In the HRN, such as that for California's restructured electricity system, the option to keep going is non-fungible, unless traded-off against grid reliability.

From Cost-of-Service Reliability to Reliability-at-Any-Price

The restructuring of the electricity sector has changed the non-fungibility of reliability from what has been variously termed, cost-of-service reliability (characteristic of the older utilities) to the reliability-at-any-price forged during the crisis induced by electricity restructuring. The logic once was that the desired level of reliability was chosen and then the integrated utilities found ways to achieve this at least cost. Reality was more complex, of course. An industry insider observed, “I wouldn’t say it was least cost. It was cost-based.” When restructuring changed the obligation to serve for the utilities, the ISO became the provider of last resort under the new market conditions. That latter mandate exposes the ISO to whatever price the market cleared. The overarching (a.k.a. non-fungible) charter of being the reliable provider of last resort even overrode the ISO’s own attempts to control price. One engineering expert explained: “When the ISO needed power for reserve operations after having gotten electricity at the price cap levels, they went to the generators and said, What do you want for your megawatts? The ISO ended up paying whatever the generators wanted, so it ended up circumventing its own price caps.”

This is not to say that cost is not an issue for the ISO. We found many instances of strategies to reduce the costs of operating. The cost savings that the ISO control room tried to realize, for example, include monitoring the prices in the bid stack and looking for uninstructed deviations in the AGC. Cost savings are sought as long as they do not create a tradeoff with reliability. Given its binding mandates, it is difficult, if not impossible, to see how the ISO could ever use cost savings as a justification for a reduced level of grid reliability. In the words of one shift manager in the control room, “This is system reliability. They tell me on the phone that it’s expensive. That’s not really my concern. I don’t spend unnecessarily. But reliability goes first with us.”

The non-fungibility helps explain why the HRN kept the lights on and was reliable in the midst of all the system volatility during the crisis. The surprise is not that it took billions to do this but that institutions actually could justify those massive sums to keep a service always on and the grid reliable. One of the world’s largest economies has become so dependent on its critical infrastructure for electricity that the costs of not paying full-time attention to it must now far exceed the gargantuan costs of ensuring its high reliability.

Paying such sums, however, turned the electricity crisis into a financial crisis rather than a reliability crisis. Mid-2001, one DOE expert argued that “the crisis for the state is happening every day, not only during blackouts. The real crisis is financial. We are buying power at high rates. To me every day is a crisis. It’s staggering.” A senior ISO engineer pointed to the far-reaching consequences of the financial destabilization: “We, the ISO, the whole system, are on the brink. ... We’ve (already) hit the iceberg, we’re taking on water fast. And we’re not getting the life boats in the water on time.” Most key financial issues remain to be resolved.

The financial crisis has made one gap in the institutional design of the California electricity system indisputable: The system—be it within the HRN, RRN or RTE—contains no mandate and principle to guide how to deal with the implications of reliability-at-any-price. Nowhere in the current design is there a compelling mandate for striking real-time tradeoffs between service reliability and the cost of maintaining that level of reliability. No one has the duty or responsibility to decide that, "If the price of the next megawatt is \$X, we will declare blackouts

rather than purchase that megawatt in order to keep the lights on." Not the ISO, not the CPUC, not the distribution utilities, in the end not even the state itself turned out to be in a position to take responsibility for turning the lights off to stop the financial crisis into which we rushed forward like moths to an open flame.

Even if such a mandate were institutionalized, there currently are no accepted principles or decision rules to guide those responsible as to when, how and why the non-fungibility should be overruled in real time. In case it needs saying, it is the "why" that has to be explained to the public and the wider RTE, an assignment which proved to be too daunting for the state even when it was not burdened with a formal mandate to step in with billions to keep the lights on. Except for dimming the lights on the state Christmas tree, it is difficult to imagine our elected officials telling the cameras that today they saved us just gobs of money by shutting off the lights across California.

There are of course other ways to limit the disruptive effects of reliability-at-any-price: most notably price caps, load conservation, new generation capacity and increasing demand-side responsiveness to price signals. We argue below and throughout this report that such alternatives are important but not sufficient guarantees as they would still require real-time non-fungibility to be explicitly incorporated in the institutional design.

Why Didn't Anyone Think About Whether Restructuring Would Lead to Higher Prices?

The most accurate answer was given to us by a legislative aide midway through the crisis: "That was one thing we failed to think about," we were told. "We are wrestling with what will happen if we don't get enough power from the market... Are we going to pay any price and outbid other players or are we going to choose not to pay \$2000 per megawatt and have blackouts. In the past reliability was the number one priority at a reasonable price, but nowadays it is unpredictable in providing the commodity." Nobody thought about rising prices because of the woefully or willfully ignorant assumption that wholesale prices for electricity were going to go down as a result of restructuring.

What then was actually thought through? The institutional design introduced via restructuring offered only one direct strategy for limiting the disruptive effects of something like reliability-at-any-price. In economic terms, under the rules then governing the market, there should have been a demand-side response. Wholesale prices would rise in theory to a point where the ISO or utilities would have decided, at the margin, to shed load rather than pay the new price—if only because ultimately the utilities, and thus the ISO, would lose creditworthiness. According to this scenario, the resulting blackouts would have reduced load and thus demand, with prices eventually falling in response, other things being equal. This scenario is what some still see on the horizon. A WSCC security coordinator argued that, "They could solve this right now if the ISO refused to buy high-priced power. Over the longer term, then, the generators would bring prices down if we weren't to buy it at high prices." Another insider doubted that the utilities would have paid the hyper-prices, if they had had the option. They would have chosen blackouts, he thought, because they simply weren't in a position to pass on the costs to the consumer.

This economic scenario was never played out because the ISO felt compelled to act as provider of last resort while the state government stepped in through the billions CERS used to procure generation during the shortages. Note the implications. The non-fungibility of reliability is an institutional feature of the California electricity system, not a formal mandate of specific organizations operating in that system. To put it differently, it is a systemic rather than organizational mandate by being a feature of the California electricity system itself rather than merely the mandate of its HRN or RRN. Thus, the notion that one can better address or attenuate this non-fungibility by redesigning elements in the HRN or RRN is at best wishful thinking or at worst a waste of taxpayers' money. As the *Financial Times* (Kay 2001) put it at the end of 2001, "No economist has ever claimed that the price mechanism can ensure that supply equals demand 24 hours a day, 365 days a year. But that is what we expect from electricity suppliers."

In economic terms, the non-fungibility of reliable electricity translates into an almost perfectly inelastic demand for that electricity. In the short term, no matter how high the price of supply, there was little or no response from demand: Load did not go down (except for voluntary conservation) and all available generation was still bought when needed, first by the ISO, later by CERS. There has been much discussion over how demand elasticity can be increased, mostly by getting price signals to the retail side, e.g., through real-time metering. We return to this issue later and argue that here too such interventions do not take away the need to incorporate the non-fungibility of reliability into institutional design.

Non-Fungibility in Real Time

Reliability, in brief, is non-fungible in real time because of the structure of the decision situation in which the ISO and others (e.g., CPUC, the state government) find themselves. At every instance, the costs of not providing the reliability are clear and immediate: blackouts or the prospect of worse. In HRT terms, casualties are known. Those casualties always outweigh the marginal costs for keeping the service going, when real-time reliability conditions of high system volatility and variable options are their most pressing.

In one sense, the situation resembles decision making under entrapment: For every incremental decision the cost of continuing is less than the cost of getting out, albeit in the end the total aggregate cost of having continued may far exceed what would have been the case at the outset (Janis, 1982). The comparison of non-fungibility and entrapment is not meant to be all negative. Indeed, we agree with HRT that non-fungibility is an important and necessary condition for high reliability. The comparison is intended to highlight one important issue, however. Over the longer term, the non-fungibility of reliability may have negative, longer-term costs that exceed the damage of incurred by periodically shedding load. For example, the state's financial crisis demonstrates how an overriding commitment to reliability in real time can end up compromising reliability over the long run. Resources used in real time during the crisis are unavailable for longer-term investments. On the other hand, the recent crisis could boost longer-term reliability. Now that the price of reliability has been made so clear, the needed investments may be accelerated, with real-time non-fungibility ending up reinforcing long-term grid and service reliability. In fact, perhaps the best way to evaluate future investments in the grid is in terms of the extent to which they are preconditions for increasing the variety of network options to respond to the bouts and seizures of market-induced volatility that certainly await us.

The longer-run is important, because it is precisely over longer time frames that grid and service reliability become more fungible and thus subject to a different dynamic (Figure 6-1). A recurring point made by our interviewees was the need for new investments in transmission and generation. An electricity expert: “The time constant for reliability is measured in cycles. The time constant for planning reliability, however, is measured in years. Reliability is real-time, but the generation and transmission systems you rely on take years to get built. That has always been a struggle and it's a task that now is much more difficult than before.”

These considerations are brought together in Figure 6-1. It is, of course, stylized, but the positive slope originating from the origin indicates that zero fungibility of grid and service reliability is most associated with real time (ignore for the moment the different slope lines for an HRN and HRO).

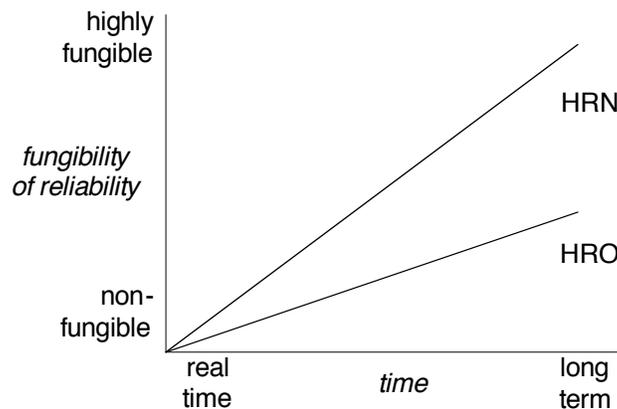


Figure 6-1
Fungibility of Reliable Electricity Over Time

Simply put, reliability is fungible with regard to long-term investments in transmission and generation and here we do find examples of tradeoffs between different types of investments producing different levels of reliability.

But there is a downside to treating all inputs as variable over the long run, even including the costs of reliability. At these longer time scales, reliability is no longer an overriding priority—to the distinct concern and considerable worry of many in the electricity sector. Investments end up being postponed, their importance debated. A New York Public Service Commission staff member explained how they faced the same issues:

All the surveys I've seen that try to value reliability at the consumer end come out very conservative. Given the system has been so reliable for so long, they do not readily have a reference for how much an outage would cost. Additionally, businesses especially lowball their estimates of how they value reliability. They figure they will be charged more for electric service if they value it higher. Unfortunately, these are the surveys that have been in use to calculate the economics of upgrading the transmission system. You really need to have tight supplies and be experiencing problems in order to get people to see the need to invest.

The restructuring of the California electricity system demonstrates that other goals may and often do dominate these longer term decisions, while their positive impact, if any, on reliability has not been assured, let alone demonstrated. Indeed, competing goals may end up dominant, precisely because their impact on real-time reliability in the future is not known.

Accordingly, we would expect a fungibility slope for a conventional HRO to be less steep than that for the HRN (Figure 6-1). HROs typically operate in more stable organizational niches that protect reliability against tradeoffs over the longer term. Thus, while we find discussion on the appropriate level of long-term, reliability-enhancing investments in the nuclear power sector, such discussions center around nowhere near the variability of investments we have witnessed in the California electricity system.

On the basis of HRT, we would predict that when casualties arising out of misjudgments and errors are not known, reliability becomes more fungible, other things being equal. When reliability has become fungible, there are fewer institutional guarantees it can be ensured thereafter. Restructuring compels long-term reliability to be part of the business cycle: “You are inevitably going to get booms and busts, which work against reliability,” in the words of one electricity grid operations and planning expert. Recent discussions have focused on measures to even out these fluctuations and maintain ongoing investments in reliability. The good news is that, in the best of worlds, these forward investments and contracts should increase price elasticity and the chance to say “no,” that is, increased fungibility and thereby reduced risks of reliability-at-any-price.

Integrating Non-Fungibility in the Institutional Design

To recapitulate: Reliability of the grid and its always-on service was de facto non-fungible in real time during the California electricity crisis and notwithstanding market options to treat it otherwise. The current institutional design has no explicit provisions that acknowledge and incorporate real-time non-fungibility of reliability. Its incorporation is important for two reasons: first, non-fungibility is a positive factor in ensuring high reliability under volatile system conditions and, second, its vital role makes explicit why and how the system can avoid producing another crisis. One engineering expert argued that,

The idea is to think about reliability in terms of risks. There are certain risks that are diversifiable and there are risks that cannot be diversified. In a market environment it makes a difference what risk you are talking about. The question you have to ask is: What is the essential reliability you want to keep. Yes, that’s the non-fungible part. For the rest, it’s just a matter of price and risk you want. If you don’t want the price or risk, you want to buy something like insurance, long-term contracts.

Because no one in the restructuring process was ever explicitly assigned the duty or responsibility to turn off the lights in response to increasing costs, the current set of mandates and arrangements implicitly institutionalize the non-fungibility of electricity as an always-on service.

Unfortunately, it is in no way obvious how changing the institutional mandates would succeed in assigning that duty or responsibility. Suppose one candidate for this responsibility—the ISO, the utilities, the CPUC or the California Power Authority (CPA)—was actually assigned a revised mandate to now tradeoff real-time reliability against other factors, especially cost. One senior state government official argued that just such a responsibility should be with the CPA:

Right now there is no one in the ISO that can make the tradeoff between reliability and the other things. This is not looked at. A proper evaluation in that sense you can do at the Power Authority... The [decision about when to go into the emergency] stages will remain in real-time operations, that will not go to the governor... ISO says, if supply, for whatever reason, is not there, we go into a Stage X. That is the reliability function and there is no difference with the way it was in the past. So those decisions on supply influenced the stages, but ISO has discretion over that.

This description clearly incorporates decisions over what is or is not fungible into a proposed institutional design. Yet it is worth repeating, can the reader imagine a public official explaining to the nightly news audience that once again they had no other way to save the state money except to throw it into darkness? If this were indeed a plausible scenario, where are those protocols that tell us where, when, how and why non-fungibility no longer takes overarching real-time priority?

Understandably, proposals to increase demand responsiveness to price have come to the fore, because they locate the duty and responsibility to turn off the lights outside the HRN and RRN, right into the RTE. A senior ISO official saw this as a logical necessity:

Someone has to make the choice to have high price electricity or shut it off. If you don't have the money, you don't have the money. Electricity is not a right. We're shifting back to the Stone Age right now. The governor now says cheap electricity is a right. The CPUC can't make that decision. They are still protecting the consumer. Choice has to be where you can shut off the load. Only the consumer can make this decision.

Real-time metering and other measures to increase responsiveness in demand—such as conservation, differentiated retail contracts with different reliability and cost tradeoffs—have received much attention. And for good reason. One ISO shift manager told us: “What happened in response to the [stage] warnings? People didn't really believe us until we started shedding load. Now people know. You see conservation, it is an issue now.” Current proposals for increasing demand elasticity are also promising.

Such measures do not, however, move the system away from reliability-at-any-price, but rather provide incentives for the “at any price” part to be less disruptive. Why? Because it would be wrong to think that increased demand elasticity would take away the non-fungibility of reliability altogether and replace it with tradeoffs between price and reliability as a quality-of-service. We have examples from other markets of where demand remains by and large inelastic even in the face of strong price signals.

Take European experiences with car mobility. The variable costs of driving have risen dramatically over the past years, yet the number of kilometers traveled also continue to increase.

Mobility research suggests that the demand for mobility is so inelastic that to get a demand response, prices would have to go up astronomically, which would never be politically feasible. The same crux holds for electricity. “A lot of discussion goes on about price, but you don’t care about price when the power goes out—public safety is at risk,” summed up an out-of-state utilities expert. Would real-time pricing actually be at market clearing levels or would there be caps to protect consumers—the latter reintroducing non-fungibility through the back door? We think the latter, if the California electricity crisis is any guide. In fact, we would expect that the cap price would be politically feasible *precisely because* it actually reflects the consumer willingness to pay. As with car mobility, politicians determine the price cap in light of what consumers would, albeit grudgingly, expend.

Furthermore, price signals lead to different responses under different conditions. What happens if during water droughts, irrigators come to believe that the real price of water is going to go up and thus respond by hoarding water they may not even need later? The hoarding in turn raises the price of water by hastening shortages even more. Or residential users of electricity in summer now read their residential meters in real time and interpret a small but new price increase as a signal that prices will continue to go up, beyond the limit of what they are willing to pay. They too start “hoarding.” That is, they think that within an hour or so, they will be facing their own mini-blackout. Consequently, they start “buffering,” for example, by turning up the freezer so that it’s colder when the prices force them to turn the refrigerator down or temporarily off. How extendable are such examples to the electricity sector? We do not know, but we do know that under any incentive structure, we have to anticipate non-linear responses—especially when it is reliable electricity rather than corn syrup with which we are dealing.

The main point here is that nobody knows just how elastic the demand will be, say, under residential metering, as well as just how that elasticity will play out in terms of volatility and reliability (i.e., we must expect household load volatility to increase at the introduction of metering). Before the crisis, we saw a stream of reports predicting that the demand for reliability would only rise even further, even to the “nine nines.” Now we see a complete about-face, where we are told that people and businesses will accept reduced reliability if they can save money that way. True, there are positive experiences with real-time pricing in some regions of the U.S., suggesting demand is indeed elastic enough to help “peak shaving.” While all these measures are promising, we should however avoid making the same mistake as in the past by not taking seriously the non-fungibility of reliability—especially with respect to the institutional design of a network that is increasingly operating and performing under real-time performance constraints. It stands to reason that the absence of such provisions will leave the system as vulnerable to volatility as it did during the crisis.

What does this mean for institutional design today? Some of our answers are left to Chapter 10 of the report. If non-fungibility of reliability is the starting point, then it seems appropriate to determine just what kinds of tradeoffs between grid and service reliability face institutional design proposals (Table 6-1). We again start with the focal organization, the ISO.

Table 6-1
Fungibility of Service and Grid Reliability

		Grid Reliability	
		<i>Fungible</i>	<i>Non-fungible</i>
Service Reliability	<i>Fungible</i>	1	2
	<i>Non-fungible</i>	3	4

By tariff, the ISO is in cell 2, where grid reliability is non-fungible while service reliability is important but ultimately of a lower priority if and when jeopardizing grid reliability. For instance, Stage 3 or a controlled blackout is declared because of an imminent path violation or other threats to the grid.

The California electricity crisis put the ISO de facto in cell 4, where the pressure to maintain service reliability has become so great that it has made service reliability non-fungible in much of real time. Officially, the ISO still maintains grid reliability as its first priority, but the increased non-fungibility of service reliability has exposed to the ISO to increased tensions between these hitherto non-competing reliability mandates. It is scarcely a coincidence that during the crisis reliability standards for operating reserves were being questioned and redefined.

Moreover, it is unclear how these tensions will play out. It is not inconceivable that it could pull the ISO towards cell 3, where the control room operators take too many chances under "just-for-now" performance conditions. A countervailing pressure is to make service reliability more fungible more often, through, e.g., the demand responsiveness programs like residential metering mentioned above. This would pull the ISO increasingly into cell 2.

But what about cell 1? Here are the ISO marketplaces for grids and services, which are at least on paper both fungible in important respects over the longer haul. It is the prospect of this cell which is truly restructuring the California electricity system, albeit in unintended ways, and where we see what may be the most compelling institutional redesign now taking place. For what we are witnessing, however ad-hoc and piece-meal these changes may be, is simply this: As long as cell 1 is where electricity markets are designed to be formally and officially, there will always exist substantial organizational and institutional pressures to move elements of those markets informally and unofficially into cells 2, 3 and 4. Cell 1 is the marketplace, and as such exhibits all those centripetal forces of market-induced volatility which push the system into real time and out of longer-run operational horizons, i.e., that push service operations into treating reliability as non-fungible in the last minute. To put it another way, as we saw in Chapter 4, deregulated

electricity markets cannot exist outside the institutional design of an HRN. The HRN must be given in order for real-time market prices to clear in the same way a distribution of income and wealth must first be treated as a given in order for any prices to clear. As long as important markets for electricity are located in the control room of a grid system operator, there will be substantial pressure to move those important market activities into cells 2, 3, or 4, particularly when real-time reliability of the always-on service and the grid is treated *throughout the system* as the one priority that cannot be traded off against other priorities.

If the preceding is correct, then we predict that there will have to be developed a much more thoroughgoing, "all-the-way-down" economic rationale and theory for the non-fungible components of service and grid reliability. There are just too many economic implications of the non-fungibility not to have a sounder economic justification for what we are observing. We already see some of the transformation in the economic rationale. Certainly, new ISO market rules and software better accommodate the realities of the grid that the ISO manages. The proposed Security Constrained Dispatch system orders bids in the real-time imbalance market not only in order of price but also in terms of their contribution to grid reliability. The new models for congestion management of the grid after the demise of the PX also trade market transparency for accuracy over the state of the physical grid. In addition, the adequate capacity requirements for load-serving entities constitute a way of getting more realistic schedules from the SCs, thereby enhancing managing for grid reliability. But other changes will have to be made, not least of which in the economic theory used to justify deregulation. These are discussed in Chapter 10 recommendations.

First, though, we need to have a much better understanding of just how the electricity markets failed during the electricity crisis and the fallbacks that have emerged to carry market activities on by other means.

7

MARKETS AND THEIR FALLBACKS

“Reliability Through Markets” was not just the ISO’s logo. It captures the core principle of the institutional and organizational design of the HRN. With generation, transmission and distribution in different hands, separated by the firewalls of hard organizational boundaries and competing objectives, a flow of transactions through the congestion, ancillary services and real-time imbalance markets was to coordinate network operations in order to provide reliable electricity. We were repeatedly struck in our ISO interviews by how many control room operators believed that the market structure could work, if given a chance and how it had more or less worked before the crisis. “Every one complains about the market being flawed,” one of our ISO interviewees told us, “but so too is the IRS. Everybody tries to find loopholes there, but no one says get rid of the income tax.” Most of us would love to get rid of the tax, but none of us thought that Houston, rather than Washington, would get the income.

During the California electricity crisis, that ISO structure broke down and the markets at the heart of this design atrophied or became dysfunctional. “Making a killing” in ISO markets became the topic of conversation in bars, backrooms and law offices of the world. “Electricity is not like other products, you cannot store it, but we’re still using it like a commodity,” lamented an ISO grid resource coordinator at one of the market desks,

Now, any other commodity in the world you can't do the same thing as is now going on in this market. I have a contract with you to produce you a jacket of a certain size and shape, but what I deliver to you is something without sleeves. You don't have to take it. It's a contract and has to be enforceable and clear. It's not as clear as it should be in the ISO. The tariff is too convoluted. There's loopholes that give people more chances to go outside the parameters they are supposed to be working within.

There were no provisions in the original tariff to deal with these eventualities. Events made clear that the market design had no analogue to the engineering reliability standards for the grid—e.g., planning for N-1 contingencies. There was no NERC for market reliability. Nobody asked about the rules and limits for market outcomes in terms of impacts on service and grid reliability. “No one thought about this happening until it actually happened” is how a WSCC security coordinator put it.

While the market design broke down, the California electricity system did not. One major surprise of the crisis is that a fairly new ISO within a never-before-tested HRN were able to reconfigure the network to now function under “Reliability Without Markets.” The market, in brief, had its fallbacks.

The aim of "just-in-case" performance, that is, "ideal normal operations," is to ensure reliability through design-based redundancy to deal with unexpected events. Not only was there no concept of "market reliability" in the design of the HRN, there were no fallbacks in the design to deal with the kind of market events we saw during the crisis and observed first-hand in the ISO control room. While the fallbacks we observed were not part of the design-based redundancy, they did emerge and proved to be not antithetical to markets. In fact, an important part of the repertoire of ISO options that surfaced during the crisis was the ability to carry on the market through different channels. If war is politics carried on by other means, then the electricity crisis was markets carried on by other means too. All kinds of market transactions between HRN partners—including those paradoxically named "out of market" (OOM) transactions—kept the electricity reliability as high as it was. We also found non-market fallbacks for the failing markets, such as informal network coordination and organizational adaptations. This ability to care and compensate for "market failure" contributed significantly to the relatively high levels of reliability observed in the crisis. Indeed, this chapter argues that the ability to institutionalize such fallback mechanisms is central to future high reliability in California's energy sector.

And Then One Day, the Markets Were Gone

We have discussed in the preceding chapters how the California electricity crisis effectively shifted ISO market transactions significantly from the day and hour ahead markets to the real-time imbalance market. But even that latter market was pauperized. "I was here, working as a new gen dispatcher, when I saw the market collapse. From one day to the other there were no more bids coming to the BEEP stack. There used to be out-of-state bids in these bids. One night I came in and a list of a few pages had become half a page without any out-of-state bids," a member of the CERS purchasing team and former ISO operator told us. The "BEEP stack" this former control room operator refers to is the real-time imbalance market, a crucial market for the ISO in maintaining reliability (Appendix A). One of the real-time grid resource coordinators, also known as "BEEPers," reported having to work with as few as 25MW in the stack to follow real-time fluctuations of the load—where that load can be anywhere between 25,000MW–51,000MW (Appendix A). Another BEEPer told us about a "really bad day. You come in, and you have only 7MW on your list in the BEEP market." An ISO shift manager observed one day how his BEEPer "used to do dispatches every 10 minutes, but now he has dispatched the whole list in the first 10 minutes."

By the time the real-time market atrophied, the other market desks had long had cardiac arrest. The day-ahead market faced a variety of chronic problems. Not all the forecasted load was scheduled in the market, leaving a major part of the energy to be scheduled at the last minute. That also meant that congestion management, which is a complicated process and must be done beforehand, now was left to real time, putting great pressure on the operators. One lead official in ISO market operations explains,

The biggest problem last year was that we had days where load is forecasted to be 42,000MW, but our scheduled resources in the morning were 32,000MW, leaving us 10,000MW short that day. How do we deal with this? Can we put a unit online? No we can't. Can we purchase energy prior to real time? No we can't. The ISO does not have the authority to commit resources prior to real time. So how can we trust the market to

deliver? The answer is, the market can deliver, but at what price? Operators end up purchasing large amounts of energy in real time to balance the system and maintain reliability. 99% of the planning has to be done prior to real time. Real time is only to react to what you missed. Real time is not ‘I’m short 10,000MW in the day ahead and I’m not doing anything.’ Most of the time things did come together, but at a very high price.

When the day-ahead market desk tried to schedule electricity during the crisis, it failed. The schedules were not a realistic prediction of the flows in real time, ultimately because “the SCs are gaming fictitious loads for congestion,” in the words of one day-ahead grid resource coordinator. The scheduling coordinators did this, in our interviewee’s opinion, to avoid congestion charges or for arbitrage purposes. The coordinator also told us of other strategic gaming behavior along the same lines involving one of the munis. The result of unrealistic schedules was that “the dispatcher cannot trust the schedule,” in the words of the senior ISO official responsible for market operations.

The day-ahead market desk was also meant to do congestion management through “adjustment bids.” But too few adjustment bids came in for it to be considered a “market.” In part this was due to the built-in obsolescence of the original market design: The PX died rather than thrived. An ISO engineering manager explains, “Congestion pricing was built around the institution of the PX, where there would be one SC for many generators and the SC would have generators on each side of the path. When PX ‘went away,’ . . . we had to go to non-economic adjustments, because there was not a very competitive congestion market.” Non-economic means, of course, it is done outside the market. A senior programmer in the ISO market operations adds that going outside the market not only makes it more difficult to deal with the market participants, but further undermines the accuracy of the forward picture of the system,

There weren’t enough adjustment bids and we are allocating transmission on pro-rata basis. Market participants don’t see the whole picture--that is, what causes the pro-rata adjustments. So from my perspective, having to explain the reason for the adjustments is troublesome.

So too in the view of an ISO engineering manager:

[This prorating strategy] has affected reliability to the extent that this means we do not have an accurate forward picture. Uneconomic congestion management solution via the software makes it look like we are reducing load when in real time it actually isn’t.

The day-ahead desk is the place not only where congestion management is to take place but also where ancillary services are to be procured. Those bids dried up as well when the shortages increased. A shift manager at the ISO told us that

...there are very few bids for ancillary services right now. The original idea was: We look at the forecasts, see what the schedules are and then get ancillary services in the day-ahead market or defer this to the hour-ahead market. There are very few bids for energy on the interties. We are often deficient procuring the replacement or regulating reserves. But there are no real specific standards there. We do need spin and non-spinning reserve, so having no bids here is a problem.

Not to put too fine a point on it, the day-ahead market operations started to drop like Icarus from the sky when the crisis heated up.

The hour-ahead market desk scarcely fared better. Although actually running three hours before real time because its operations are too complicated to be completed earlier, it is basically meant to adjust the day-ahead schedules to the actual load during the day, as well as make further adjustments in ancillary services. Its problems were similar to those of the day-ahead market. We happened to meet an ISO hour-ahead grid resource coordinator at an exceptional time: “If you’re able to successfully meet your requirements in the hour ahead market, that’s a good day... I did that tonight for the first time since the end of 1999.” We were talking to the operator on June 1, 2001. A bit later he added, “Anyway, you know we don’t have functioning markets here.” One of his colleagues noticed the loss of functionality this way: “There’s probably half the number of schedules and half the amount of megawatts. There is not a lot to get fixed.”

Whatever markets were there in name, they were no longer markets in the way originally planned. A senior market operations manager described a bad day during the market failures:

[An] example of a bad day is when we have congestion, but it cannot be resolved using economic bids submitted by the market participants. In addition, the ancillary services bids are not sufficient to meet the requirements and we lack capacity to meet the real-time imbalance energy requirements. These conditions have the real potential to reduce grid reliability.

The financial crisis that hastened the fall of the markets took care of the rest. “The destruction of the utilities’ creditworthiness and the resulting responses by suppliers shattered all vestiges of a normal market,” is how Mensah-Bonsu and Oren characterized the situation (2001, p. 2). This neatly summarizes much what we saw during mid-2001.

Markets on Life Support

The California electricity crisis made visible the massive intensive-care infrastructure that surrounds the markets and makes their existence possible. A repeated theme of this report is that markets draw their very lifeblood from being in an HRN whose control rooms and wraparounds take grid and service reliability seriously. The market is not a virtual place where supply and demand intersect and prices clear. It is a patient kept alive by all that science and technology have to offer.

Less metaphorically, the ISO markets are a bricolage of software programs on servers through which market parties can articulate, adjust, accept bids, prices and charges are set, markets are cleared and, ultimately, all these transactions are transformed into actual electricity transfers. These programs are designed, managed and redesigned continuously, as is the way they are operated. As long as the wholesale markets operate smoothly, the organizational and technical infrastructure around them stays invisible. During the crisis, however, the markets were put on highly-visible life support to try to get them back to “normal functioning,” or at least preventing them from withering away altogether. “Last night, I struggled with the software for seven hours.

One of the modules were incorrect. It's so massive, all the software we must use," said one of the hour-ahead market desk coordinators.

At least as important, if not more so, markets were put on life support to forestall any major grid disruptions and collapse. Managing the discrepancy between the (dys)functioning markets and the requirements of the technology became an overriding concern because of restructuring, as a WSCC security coordinator sketches,

NERC and CPUC never envisioned these kinds of problems, for example with price caps. We tried to make the market work. We bent over backwards to give the markets room. A problem would prop up and we'd put a bandage around it and then there came more leaks. We set the \$750 caps in place to help last year and what happened? More problems, as generators figured out that they could hold back electricity to the last minute and get \$1500 instead. These software issues and market functioning issues weren't envisioned when they came up. For example, last year when things were going to hell, we were short 10,000MW a day and decided that instead of waiting for the real-time market, we would try to buy on the day-ahead market. That backfired. Everyone held back.

When the bids were no longer coming in, operators could not just sit back and rely on the software and routine market procedures as before. They now got on the phone, sent out emails to market parties, issued a high price—all to “stimulate” or “excite” the market. “There are also other ways when we are desperate,” said an ISO hour-ahead grid resource coordinator. “When we are really running short, I can always send out an email as an alert.” In addition, the schedulers and later the CERS employees would be alerted and they would go out and “hunt for megawatts.” A very senior ISO official said told us that “these guys here [in the ISO control room] know every megawatt in the west and who's got it.” When an important part of the market moved into real time, they brought in a portable console and had three schedulers working to support the overloaded BEEPer.

ISO control room operators had to adapt to the volatility of the market and more and more adaptations brought them outside the original market design. “[In the past] going out-of-market was unheard of. Now everything is considered out-of-market,” explained an ISO shift manager. “Out-of-market means I have no control over it. It is a transaction between the generator and the state. It is not dispatchable to us. It is dispatchable before the hour, not during the hour.” A pivotal adaptation centered around the demise of the PX, which used to submit to the ISO the schedules for some 80% of the generation. Suddenly, ISO operators were confronted with “What do we do, now that no one sends in the schedules?,” a senior engineer in the ISO remembers. “We had to change the rules and create new rules. What we relied on were the business contingency plans done for Y2K.” The overnight transition to a system without the PX went relatively smooth because of the support and buffering of market and engineering staff in the ISO wraparound (Chapter 3).

The California electricity crisis compelled elements in the ISO wraparound to adopt a "just-in-time" performance approach that could be called “rapid-response market engineering.” It was equivalent to rebuilding the tanker while at sea, with whatever is hand and without the option of returning to port. All ISO markets were having problems, requiring constant redesign. Filings to FERC had to be prepared. Defying conventional development cycles, new software had to

designed and redesigned at fast-thinking speeds. Since many market rules are embedded in the software, responses to market failures immediately entailed software redesign. Next to market operations, information technology was the fastest growing part of the ISO.

The more-rapid-than-expected morphing of software created its own operational challenges for real-time reliability. Whereas HROs, such as the earlier integrated utilities, test new systems extensively offline and on a small scale before bringing them online into primary operations, the ISO was pushed to use operations as the testing ground for new systems or untried system elements. Without exception, operators reported fighting with software glitches. “The changes in the software were never thoroughly tested because of the hurry mode to change the market,” describes an hour-ahead grid resource operator. Software glitches were among the top causes for “bad days,” right next to lack of resources and SCs not following dispatch instructions.

Without the wraparound’s intensive care of market and grid operations, the markets would have imploded. Major coordination mechanisms of the system faltered, but the ISO and others in the HRN managed to keep the patient in one piece without getting rid of the market design completely so as to maintain service and grid reliability in real time. Here is how a senior market operations manager described the result,

I’m encouraged that there are many things in the market that haven’t unravelled when they could have. For the most part, right or wrong, for good or bad, our market has not changed significantly. It has been tested, taxed, we were taken beyond what the market could provide. But the crisis has not gotten rid of the markets altogether and forced us back to command and control. Since that hasn’t happened, yet, and I believe we are on the backside of the crisis, we can get clearer heads and get the correct balance in a logical way.

This conclusion is only possible because there were fallback mechanisms in place or improvised that filled the coordination gaps. The next section outlines three of the mechanisms that carried the market on by other means.

Rebuilding the Links: Fallbacks for the Market Design

Since electricity was provided reliably, even though the ISO markets were dysfunctional, perhaps the markets were not that important to operations after all? Just throw the state treasury at the problem and. . .*voila!* Anyone observing the ISO control room during the crisis would, however, have concluded otherwise, as it took much hard work and markets by other means to keep the lights on, and even then with no guarantees and little to spare.

How so? Comparing experiences between ISO and PG&E, one of the ISO market operations manager told us,

It is different from the older centralized world in PG&E. There are significant differences between the way operators now control the system with the markets and how the system was then operated. The operators were in complete control then. Now the market requires more work for the operators to control the system. We are in the middle of trying to find

balance between operations and markets. There is no doubt that operators have learned how to rely on markets to operate.

A WSCC security coordinator, also comparing past and present, illustrated the reliance on markets but in more negative terms: “When I was shift manager under the old system, I used to call and the generators would do what I told them. You were one step below God as shift manager. Now the generator just ignores you.” Private generators do not ignore the market, but what happens when they stop functioning as needed by the HRN’s focal organization?

Three basic fallback mechanisms that stretched across the gaps in the network left by the atrophying markets: informal networks and coordination, contingent command and control, and organizational adaptation. Before turning to each, it is important to highlight that the fallbacks are not antithetical to “market solutions.” As we will see, all use, rely on or at least allow for market transactions. The mechanisms, however, do fall outside the formal design for restructuring. In that sense, they were never part of deregulation, but are emergent properties of the restructured system. As the earlier interviewee put it, the same institutional design basically remained in place throughout the crisis. In this way, the fallback mechanisms have not so much replaced the markets as have filled niches and opportunities created by that market restructuring. The significance of this last point is that the emergent fallbacks were functioning, viable and effective even when attempts to redesign the formal markets met with fierce opposition and failed to prevent the official market structure from collapse.

Informal Networks

By this point, it comes as no surprise to the reader that informal networks are a major reason why market behavior carried on even though the formal markets themselves did not. We have already seen that informal networks are crucial to the operation of the wraparound (Chapter 3) and the always-on management of electricity (Chapter 4), both of which are in turn at the heart of the rationale and operation of markets and their fallbacks.

In the paper mache world of AB1890 and regulatory documentation, the flow of market transactions do not require operational communication outside the market software and official tariffs. The tariff actually prohibits many of these communications. During the crisis, as we have already seen, rich networks of informal communication and coordination emerged or resurfaced to play a determining role in securing real-time reliability under performance conditions of high volatility.

The process of outage coordination provides a telling example. Outage coordination covers generation and transmission. For generation, the ISO's official tariff only required RMR units to get permission for outages. The remaining ~80% of generation capacity submitted an annual plan. Officially, that was the last time generators had to contact the ISO. Unofficially, they still phoned in, according to the manager of outage coordination, because that was what traditionally was done. The lack of ISO authority to deny or unilaterally reschedule planned outages had not been a major problem until December 2000, when the ISO suddenly faced emergencies precisely because of all the outages, hitherto an unprecedented event. An ISO shift manager observed,

“For flexibility we need to get [the generators] back on line. The market has to get back [to working].”

In the absence of formal authority to coordinate outages, informal means came to the surface, given the overarching reliability mandate of the ISO. Much revolved around informal give and take, as the manager of outage coordination explains,

We are pretty successful in coordinating outages of different units within one company. We tell them we can only have one outage at a time, not everything at the same time, for example. But they can choose when to take what offline, and there is some give and take here. Another example is where they have an RMR unit and other non-RMR units. For the RMR they have to get permission under the contract, not for the others. The RMR would give us some leverage over the other units. We were more flexible with the RMR unit if they would agree to schedule the outages for the other units at more opportune times.

The back and forth is not so much opportunistic as it is a ongoing process management that rewards cooperative behavior. Generators and transmission owners who submitted their plans on time and took them seriously received more flexible coordination. To do so requires that the coordinators function as an institutional memory of the behavior of different generation or transmission owners. “You have a memory and relationship with the different generators,” the ISO outage official told us, “and that factors into outage coordination. You will say ‘Remember I held out [X generator] for a week,’ so those kinds of trades do go on.”

Flexibility is at a premium under real-time performance conditions for the generation owners as well, the ISO outage official continued:

One thing we are giving them is when they want the outages. We can be flexible with the RMR contracts. We cannot make RMR cover another outage, but what we can do is that when a unit wants to come off for work, we can sometimes let them come off a day early or so. That may mean that they don’t have to pay premium to get their crew in on Sunday. So we try to get them off on early or late Thursday, instead of Friday. And if you can come off early, we can extend outages also when needed. But if there is a company that takes advantage of us, we don’t do that.

We identify many such examples of informal flexibility and coordination around other tasks throughout this report.

For some readers, such give and take sounds very much like the way things were done in the older, integrated utilities. “But there is nothing new about [such behavior],” a senior CPUC planner told us,

That happened in control rooms before restructuring. . . Those kinds of informal contacts were from utility to utility and even across the border as well, with people in BPA and their markets (although we did not call them markets then). They had the same informal relationships: deals were struck in the crunch, with nothing written, nothing formal. Just on the phone: “We need energy, and we pay you this. Done.”

True, but our point in this report is that the shifts to increased volatility and real-time operations have dramatically intensified and transformed the importance of these informal networks, *especially for the operations related to the markets under restructuring*. “The market is so volatile,” an energy trader and former utility executive told us, “it is way too volatile, you try to sell someone electricity and two hours later everything’s changed. In a real market, guys can hedge, but we don’t know how to hedge in this electricity market.” Any wonder the ISO sought to develop fallback options in response?

Other ISO operators continue the story of how important informal network coordination is for market operations. An hour-ahead grid resource coordinator said: “There are different customers. You have quick and slow people. You can categorize them. Some companies have some reputations and some people have reputations. I get a sense of that.” Another grid resource coordinator described how some SCs do not know the procedures. “I wind up putting in their schedules. If I do that, they might help me and that helps California, but I’m sort of doing their job for them.” A BEEPer explained how “When you get to know them over time, there are some that have more expertise, and there’re others that are less skilled.”

Some ISO staff, e.g., schedulers, had little need for informal communication and coordination the first two years after the market started. The market took care of it. But during the crisis, they started calling generators more and more, going out on the market themselves. “Schedulers are the backstop positive redundancy,” is how a senior ISO scheduling manager summarized their informal role.

When push really came to shove, and ISO operators were hanging on under “just-for-now” conditions, informal communication became mandatory. These times were likened to combat situations by our interviewees: Peoples lives are at stake when traffic lights go off or utility crews are scrambling to just keep the lights on. A senior executive of an energy company told us how at these times he would get calls from his counterpart at the ISO or he would call them himself (Chapter 4). Such out-of-channel calls are necessary and indeed becomes probable, because all participants in the HRN know when the margins are very tight and the ISO is running out of options. We saw the same sense of informal network coordination behind the description of how, during the demise of the PX, the “generators played ball and everybody cut us some slack that week. They could have played hard ball, but they didn’t.”

A phone call to exchange information is crucial when it helps the ISO get a better view of available generation and transmission capacity. Several interviewees mentioned that many generators have reserves to cover their commitments, but that reserve capacity is not visible to the ISO and has not been counted in the official reserves. “Say we know a 750MW generator is out there,” a WSCC security coordinator explains,

It is actually generating 500MW and has put in a bid of 100MW for non-spinning generation. So there is 150 MW that is not accounted for but which I know exists. It’s really there. What’s more, this uncounted capacity is actually used in case of major system disturbances. In these disturbances, the governors on the generators are going to automatically react or respond to bring frequency back into line.

Under the highly volatile conditions of "just-for-now" performance, such information can make the difference between going into blackouts or hanging on.

These examples are not meant to suggest that informal coordination always replaces formal design or always works when required. When an ISO shift manager called the owner of the generation unit to inform him that his actions were threatening system reliability, the owner was more worried about how he would be compensated financially before doing anything. Another time, a generator was asked whether maintenance could be finished before the summer. The answer was "no" — unless, of course, the ISO or someone else wanted to pick up the \$25 million price tag. Informal networks could not prevent the California electricity crisis or even resolve it. But they were instrumental in adapting to changing performance conditions that were never designed for, let alone envisioned, through the tariff and other formal design arrangements. The crisis induced by restructuring set service and grid reliability against each other and it was only the informal that bridged the gap up until something more could be done.

Contingent Command and Control

Many system insiders observed the gradual return of "command and control" provisions for the ISO during the crisis. A WSCC security coordinator told us: "With the current crisis, there's been a swing back to reliability and that's good. If market entities were required to follow the tariff, we wouldn't have these serious problems... If you want reliability, there has to be an authority." A senior ISO engineer concurred, arguing that generators who were not following dispatch instructions were a "major problem." We heard similar views from the RRN. "Unless we have control over the plants, there's nothing we can do," a senior CPUC staff member argued. A senior market operations manager at the ISO told us,

I think in hindsight the market rules were probably designed too much from the marketers' commercial perspective and not balanced by the operational realities of maintaining the transmission grid. The pendulum has started to swing back towards the physical realities of the system. How far it is going to swing back is an open question. If we're careful, it could swing to a balanced point. However, if we are not careful, it could swing back to a complete command and control world, in which little or now market opportunities exist.

Yet the notion that the pendulum is swinging back from markets to command and control is somewhat misleading. The "command and control" mechanisms that filled the vacuum created by the disabled market design did not replace those markets, but became available only when the markets ceased to ensure service reliability and came to threaten grid reliability. It is better to see some fallback mechanism as contingent command and control, that is, provisional both on the ISO's ultimate ability to declare a Stage 3 emergency and on the high system volatility which necessitated improvisation and fallback alternatives to distressed markets in the absence of a Stage 3. In fact, the emerging arrangements discussed in this chapter for operating markets by other means were possible, we believe, precisely because the ISO chose *not* to exercise the option it has to declare a Stage 3. If declaring a Stage 3 is thought of as the ISO's ultimate fallback option, its other fallback market mechanisms under real-time reliability performance conditions can also be thought of as contingent on the space created by not exercising that option

even when it could under the conditions now facing the ISO. Thus, rather than moving the system back to pre-restructuring conditions, fallback arrangements were created precisely because restructuring “kept in place” conditions under which ISO command and control is asserted through a Stage 3 declaration as well as “put into place” new conditions which required alternative options to the original market design. As an ISO market operations official put it, “I am not saying that we should go back to command and control, but we do need checks and balances. . .”

The contingent nature of these “checks and balances” varied in form and intensity throughout the crisis. First and foremost, there were increased calls for better planning and authority in the forward markets. Related to increased calls was the increased use of contracts, including those the ISO has for RMR units and “free voltage support” from generators, which in effect were market-based ways to carry on command and control by other means. Legal agreements also figured in another way. A shift manager at the ISO described: “[The system] has been going in that direction [of command and control] for a while. Look at all the lawsuits and injunctions we have been seeking to force generators to keep supplying power. If it wasn’t bid into the market, we’d ordered it online. Everything was online. Now we still do some of that, but not ordering everything online.”

Note that other reason why the pendulum metaphor is misleading: the use of lawsuits, injunctions and contracts in order to get members of the HRN to treat reliability seriously. Many of the ISO’s fallback arrangements simply do not mimic the same kind of command and control that was in place before restructuring and its crisis. In some cases, the fallbacks work less on “command and control” through the activation of vertical authority and organizational hierarchy than through the use or threat of legal action to determine who has “command of control” under the circumstances in question. The use of the court system recognizes the fact that legally independent organizations are involved in interdependent specifics. This kind of “command of control” is per definition contingent on the particulars of the case and is adjudicated laterally among equals rather than decreed top-down across cases among unequals. The ISO has entered into a variety of court cases, regulatory hearings, contracts, and other activities on specific problems it was dealing with (e.g., uninstructed deviations from contracted energy schedules) to find out what kind of formal authority, if any, it has over certain operations and how its command of control might be extended within the current legislation—both issues that remained unclear after the restructuring. In short, while market fallbacks were emergent, not all depended on informal networks and coordination. Some emerged very formally indeed, as perhaps best demonstrated by the many amendments to the ISO’s tariff.

To put it more broadly, the California electricity crisis has led to a bottom-up, fragmented learning process of defining and developing contingent “command and control” in the high volatility space crated between the collapse of the markets and a Stage 3 declaration. We have yet to understand fully the composite fallback mechanisms that emerge (self-organize?) from these bottom-up specifics. An electricity grid operations and planning expert we interviewed argued for a “a hybrid system of central guidance—not control—to have a complementary role to assist market. We’ve really never done a study comparing deregulation, hybrid and regulated alternatives to see which one is better for all stakeholders.” While “assisting the market” does not capture the importance of such provisions—after all, a key fallback transmission line or

generator does not “assist” the primary unit, but rather is part and parcel of its reliability—we share his conclusion and suggest that future improvements start with these contingent forms of command and control that emerged during the restructuring-induced crisis.

Organizational Adaptation

The third fallback mechanism we found was organizational adaptation. The hobbling of the grid's considerable horsepower by what restructuring called the “unbundling of generation, transmission and distribution,” was actually untethered in large part through organizational means. Most visible was the creation of CERS (the California Energy Resources Scheduling unit), part of the state's Department of Water Resources (DWR). Its creation by the state to buy electricity on behalf of the distribution utilities made it a market party with a public task: a contractor of generation and with it *de facto* load, all to be performed inside the ISO control room, which was originally to operate under a mandate that separated such outside scheduling coordinators from ISO operators. This development, understandably, raised eyebrows on the part of the operators, and a few even argued CERS made their jobs more difficult. A WSCC security coordinator in the control room argued:

Has reliability increased with the DWR in the control room? No, it has decreased. It's added another layer of hierarchy to dispatching. We already had another layer and now we have generators insisting that when they are called up by the ISO gen dispatchers, they have to get personal confirmation from DWR. Why? Generators only want to deal with a creditworthy entity. I don't understand that. They are going to get paid.

Many of our control room interviewees found it difficult to deal with the presence of a market party in the control room. “What would you think if they gave a workstation in the control room to Enron?,” one control room operator asked.

Notwithstanding these tensions, most saw CERS as enhancing reliability, given that the distribution utilities and ISO itself had problems with their creditworthiness. Thus, the ISO, as indeed did others in the HRN, learned to work with this organizational adaptation, as an ISO shift manager explained:

Now CERS is here and the rules with them are renegotiated all the time. I try to work with them, they are in it just like we are. It's going to take two, three years before the big mess is solved.

While CERS was a very visible adaptation, in many ways the ISO went even further in adapting itself to fill the vacuum left by the poverty of market design. A state government official told us:

All of these problems moved the ISO into the situation that it is now creating a demand bidding program, which are subsidies really. Of course, somebody should do this, but not ISO. Also the ISO is involved in buying generation, peaking generation. The super peaks (40-48,000MW) are only for a couple of hundred hours and should be funded outside the existing market. You need other mechanisms than the market, or you'll end having very high prices during those hours. So ISO argued rightly that you have to have different peak product. It is not the ISO's job, but the argument was right. So the ISO was pushed

into load management, generation business, it was almost pushed into the forward contracting business—that is now CERS—and it was not supposed to either. It was pushed into all this, into the pieces of the former utilities business.

The ultimate transformation was, of course, the ISO transforming itself into the provider of last resort. While stakeholder, regulatory and legislative processes took months, the ISO built the necessary organizational infrastructure because no other entity was in the position to do so. The fact that the ISO was able to adapt to the new tasks and challenges suggests that the non-fungibility of electricity over the short term relates not only to price, but to authority as well. During the crisis, the ISO took on activities it was never intended to assume, but was reluctantly allowed to because the consequences of not doing so were intolerable. Restructuring had created a status quo that was untenable, and the ISO adapted accordingly.

If we are correct in arguing the incomplete network design will persist and that highly volatile performance conditions will undoubtedly recur in the future, then the ability of the ISO to act as the provider of last resort is an incredibly valuable fallback asset. As one of the ISO market managers concluded,

I think the ISO is going to change in the future. The changes that we want to make may put us into a more active oversight role within the market compared to the initial relatively passive design under deregulation, where the ISO really only focuses on system operations.

Thus, the final irony: The failure of its markets has made the ISO much more of a market participant and in many more ways than ever imagined by its own tariff.

Conclusion

The starvation of the ISO markets and the intensive care they required afterwards not only point to design flaws in the markets themselves, but also in the wider institutional design. There was no concept of market reliability analogue to the engineering reliability standards. There were no formal provisions to deal with market failure, other than formal redesign through the regulatory route. Any formal design mandating high reliability is doomed if it leaves no space for emerging, contingent and provisional fallback mechanisms such as those discussed above. You do not design organizational cultures; they emerge when they are needed to nurture and meet formal mandates. Just as regulators increasingly understand the need for performance-based rather than blueprinted standards, so too should everyone in the RRN and RTE understand that performance-based organizational cultures have a fundamentally necessary role to play when formal misdesigns blessedly die but original mandates do not. For any system operator working within a deregulated setting for reliability management, market fallbacks have to be part of its options set. This includes its ability to shift and adapt across different performance conditions, whether formally or informally, in the name of meeting the overarching reliability requirement of balancing load and generation. In the next chapter, we turn to how the reliability standards themselves have evolved and adapted during the crisis.

A key implication of the findings is that fallback mechanisms must be further explored and retained as capabilities for the reliability of the California electricity system. To repeat the point made earlier in the report, it is not realistic, if it ever was, to plan from the assumption that once the system has been stabilized, it will not be exposed to unexpected disruptions and volatility. The literature on organizational reliability clearly insists that high reliability depends upon multiple and varied means and options to respond to the unexpected. The fallback mechanisms discussed in this chapter, or their functional equivalents, provide just such responses.

8

(RE)DEFINING RELIABILITY

One visit the control room sign was there; the next visit we saw just holes in the wall. Where did “Reliability Through Markets” go? The sign, whose disappearance was reported to us in four versions, pretty much sums up the state of reliability in California’s electricity system. It’s changing, and for many different reasons detailed in the report.

Simply put, restructuring broke up the happy marriage of grid and service reliability. Treating reliability as non-fungible and letting markets atrophy must be understood first and foremost as options to maintain grid and service reliability in tandem, precisely when the latter has the greatest chance of threatening the former under persisting real-time performance conditions. Of course, there always has been a trade-off between trying to keep the service always-on and in doing so jeopardizing the grid as a physical infrastructure. But prior to restructuring, grid reliability meant nothing less than service reliability, and there was really no ongoing worry that service would itself undermine grid. Now there is. Perhaps nowhere else is the impact of the wedge between grid and service reliability better observed than in the changing definition, standards and criteria for what constitute reliability, the subject of this chapter.

Adapting Reliability Standards to New Conditions

Our interviewees tell us that there is not one official reliability standard that has not been pushed to its limits and beyond in the California electricity crisis. The public impression is that not meeting these standards means that reliability has gone down. This obscures, however, that the standards themselves are under divergent pressures within the California electricity system—that is, reliability is being differentiated and redefined.

Reliability seems at first glance to be defined fairly coherently and consistently. There is a set of mandates from the RTE that guide organizations in the RRN to define standards that are then implemented in the HRN. The restructuring has put divergent forces at work on these standards; and the crisis has exacerbated this (Figure 8-1).

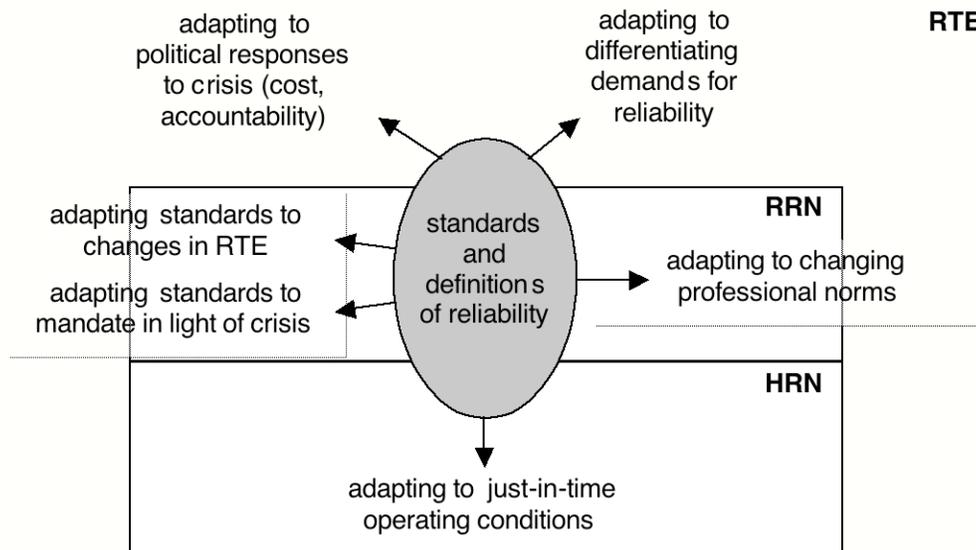
Redefinition in the RTE

For the RTE public and economy, reliability is first and foremost keeping the lights on and electrons flowing in the right direction. The prevailing sentiment is “Reliability is I don’t lose the lights,” as one ISO interviewee phrased it. What this means practically is changing in several ways. First, the market and transaction costs of keeping the lights on have reinforced and amplified thinking about what level of reliability for specific types of consumer. This drive to

differentiate the reliability of electricity by different categories of consumers (e.g., Silicon Valley industrial consumers need a higher degree of service reliability than do residential consumers) is a fundamental redefinition. A trend already before the crisis, differentiation is an increasingly salient response to the threats to reliability post-restructuring.

Second, the massive costs of the crisis underscore an important amendment to the notion of reliability as “keeping the lights on.” Not only “[is] electricity a right..., it’s a right to have it at a reasonable cost,” in the words of a senior state governmental official. Clearly, the reasonable cost proviso redefines what it means to be reliable for the RRN and HRN and is in tension with the redefinitions (i.e., reliability at any cost) taking place there.

Third, the crisis has triggered interventions from the RTE aimed to increase the control over and accountability of the California electricity system. Expanding the already populated RRN with the California Power Authority is the clearest example of this strategy. Investigations into generator behavior and associated lawsuits further expanded the notion of how parties in the



HRN are expected to operate in light of reliability requirements.

Figure 8-1
Divergent Forces Redefining Reliability

Redefinition in the RRN

The RRN, too, is adapting its definitions of reliability, albeit in often different directions. The tensions between FERC and the CPUC, between the CPUC and the CEC, and between the WSCC and the ISO with respect to ensuring the reliable provision of electricity have been manifold (Appendix A). FERC itself provides a good illustration of the tensions put into play within the RRN as a result of deregulation. How does FERC regulate?, we asked one legislative

aide. “Only in a very light-handed way. FERC has gone from a cost-of-service based approach for regulating prices to market-based rates. With some minor exceptions, FERC has said that whatever market prices are, they are just and reasonable.” “Nobody took responsibility for prices [during the crisis], not FERC,” concluded an electricity advisor to the Governor’s office.

Restructuring federalized the California grid, shifting the balance of responsibilities and oversight between FERC and the CPUC, without however identifying compelling, close regulatory oversight for grid transmission planning and investment. One CPUC insider observed,

The ISO did have an interest [in grid maintenance], but it had neither resources or authority nor institutional structure for inspection. . . We did maintenance checks. In fact, the San Francisco 1998 storms, when the PG&E system failed for days, made clear maintenance was needed. We conducted an investigation but could find few causal links. Nonetheless, it was clear to us that the PG&E system was run down. So we started to get specific about the maintenance standards and rules. But we never had the resources for spot checking this in the field.

The story about CPUC oversight and responsibility with respect to grid reliability is picked up by the legislative aide,

The CPUC is in a kind of struggle with providing reliability. They still regulate the utilities, but they only regulate the distribution service and part of the transmission service. But private parties own a lot of generation and CPUC has no economic control of these generators. Over the last months, we have had unplanned outages, due to maintenance problems, accidents, and a lot of people suggest that private companies were withholding power to raise prices up. CPUC really doesn’t have clear mandate to follow that and sanction it. Their ability to ensure reliability has diminished as well as the utilities’ ability to maintain reliability as generation has been privatized.

The crisis has brought out many such discrepancies between the reliability standards and the mandates of agencies. Many agencies are now trying to change the standards to better fit and protect their mandate and to adapt to the new demands each faces from the RTE. Needless to say, these attempt at redefinitions are often divergent rather convergent.

The self-regulation of the sector through NERC and the WSCC have their own dynamic of redefinition. An important process since restructuring has been to adapt the standards to the changing professional norms in the sector itself in light of the market conditions under which it now operates. Prior to and during the crisis, for example, the ISO has been working under much more stringent enforcement rules of the sector. The drive toward enforcement made sense under restructuring, where the old-boys networks were to be replaced by more transparent and accountable market transactions and relationships. For the ISO, the new enforcement rules from the RRN have led to some unhappy encounters. The ISO ended up paying large penalties for not meeting the new enforceable reliability standards, even though it was clear that the non-compliance was outside the ISO’s control. As a senior ISO engineer quoted earlier argued: “Are these penalties helping me to do a better job? No. Is it helping us learn to operate better? No.”

Redefinition in the HRN

It is clear that the divergent processes of redefinitions in the RTE and RRN also affect the HRN. are divergent with respect to each other and with many of the demands faced in the HRN.

Whether it is not meeting new enforceable reliability standards and end up paying fines that aren't helping anyone, or it is the senior ISO official shocked by the fact that "the governor now says cheap electricity is a right," we found many more instances where the ISO and others in the HRN are grappling with the redefinitions going on elsewhere.

For the ISO, its most important drive to redefine reliability has been to deal with the fact that during the crisis all its reliability standards were being pushed to their limits and beyond. Such breaching of standards looks to be prime facie a decline in reliability. Certainly there is the public impression that reliability has gone down as a result of restructuring. Our report's framework and argument run counter to this view.

What actually happened was not so much corroding reliability as rethinking reliability in light of the changes in performance conditions under which the always-on service is to be maintained. Instead of the older cost-of-service reliability with high operating reserves, California restructuring fractured the means to achieve reliability, in this instance, through "just-in-case," "just-in-time," "just-for-now" and "just-this-way" performance modes. The older standards were breached at times precisely to keep the lights on, with new standards evolving or emerging in the process. The effect has been to redefine both what it takes to keep the grid reliable, whatever the level of service demanded of it, and what it takes to keep service reliable without jeopardizing the grid in the process. The following sections explore these processes of redefinition in the ISO and HRN in more detail.

Adapting Reliability Standards to Real-Time Conditions

The reliability criteria and standards at issue are many. Most important, operating reserve limits have not only been questioned, but the system has operated reliably—in particular, peak load demands have been met continuously and safely—at lower reserve levels than officially mandated in WSCC standards. CPS2 criteria have been disputed and efforts are underway to change them. There has been mounting pressure to empirically justify standards that were formulated ex cathedra in earlier periods, which have since become "best operating practices." One senior ISO engineer told us,

Disturbance control standards (DCS) says that if I have a hit, a unit goes offline, ACE goes up, I have to recover within ten minutes. Theory was that during that time you were exposed to certain system problems. But who said 10 minutes? God? An engineering study? What are the odds that another unit will go offline? One in a million? So now with WSCC we have turned the period into 15 minutes, because the chance of another [unit] going offline is low, as we know from study of historical records. . . .

All performance standards are being opened up. The fact that you have to pay for it is why they are being looked at. The fact that the standard "felt good" is not good enough any more. And some will say, dammit it always worked and you are eroding reliability

now. And they are probably right, but eroding it by how much? Probably not by much and that little bit of reliability is bought at a high price. So it is an acceptable risk and it is even desirable to change it. It is no longer reliability at any cost.

As operators learn more about their grid and service reliability under "just-in-time" and other extenuating performance conditions, criteria for reliability can and must be rethought, if we expect the ISO specifically and the HRN generally to be any kind of learning organization (Senge, 1990). A WSCC security coordinator describes what he sees at issue in the learning,

These alterations don't change the basics, but some parameters were a little too 'tight' or restrictive. [Revising parameters] allows them more flexibility to recover from disturbances. It's realizing that they [the operators] are doing the best as they can. Some of the original parameters were unattainable. Other control areas had the same problem.... It's a NERC regulation that every generator that is online is supposed to be under AGC. ISO doesn't operate like that... Why? Because it is not cost-effective. . .

Operators want, in the words of the ISO senior engineer, "a bar they can jump."

It could be argued that all the rewriting of criteria refines standards rather than redefines reliability. Obviously, NERC and WSCC standards have incrementally evolved in light of engineering lessons learned and experience had. Our argument, however, is that redefinition, both in terms of grid reliability and in terms of service reliability, is actually underway. Reliability is being redefined in terms of a shift from a focus on N-2 contingencies to the single largest possible failure in generation. Shedding load was considered anathema to control room operators in the past, while now controlled blackouts and recovery are part and parcel of what it means to be reliable in the face of high volatility. NERC standards, which assume service reliability is coterminous with grid reliability, are under wider redefinition pressure as well. According to a university engineer familiar with electricity restructuring and operations,

The standards are changing, for example, for the amount of spinning reserves. NERC has stringent criteria on spinning reserves. New England has been experimenting to mitigate some of the price spikes in spinning reserves. The ISO there has constructed a hypothetical demand function (which means buying less if prices go up) to dampen price spikes. This worries NERC who wants the ISO to strictly maintain their standards. The reason is that we need X amount of reserves in NERC's view. Nobody considered that you can meet the same standard on average so that you can be price responsive and can vary your spinning reserve within a certain range. A lot of these rules will have to change.

Clearly, the tension between the older cost-of-service reliability—"we chose the standard of reliability for our customers and then estimated how to do that in least-cost terms," according to a utility director of a maintenance and construction division—and the crisis's service reliability-at-any-price is creating pressures for redefinition over just what is meant by "high reliability," especially when such tensions reflect the real-time trade-off between grid and service reliability.

The following sections describe the magnitude and direction of the redefinition of grid and service reliability taking place as a result of the electricity restructuring, its crisis and the processes underlying both. We first discuss redefinition in terms of the changes in what

reliability is taken to be. Second, redefinition is discussed in terms of changes in where reliability is taking place. Third, we ask again the question of whether reliability has gone down.

The chapter's take-home message can be stated succinctly. One principal impact of restructuring and the crisis has been to introduce multiple and changing criteria for assessing grid and service reliability and the now quite palpable trade-off between the two in real time. Accordingly, the performance record for achieving the real-time balancing of load and generation will inevitably be mixed, never a complete success in following standards but then again never a complete failure in terms of the results. Breaching a WSCC standard will be countered by having nonetheless made the balance "just-in-time," and operators can be expected to see mounting examples of the divide between standards of reliability and the actual production of always-on electricity. In the operational world of mixed, real-time performance, hard-and-fast standards matter less than "using your best judgment" and "judging each hourly balance on its own merits" in order to walk the fine line between service and grid reliability. Or to put it differently, in order to preserve reliability, the bandwidth of permissible performance by operators is being *de facto* widened albeit not *de jure*, thereby more and more accepting the very real trade-off between service and grid reliability.

The mixed performance record will however be read wrongly by regulators and politicians in the RRN and the RTE who will see the record as an excuse to redesign the HRN. They will call for more redesigns, more regulations, more demands that "the ISO follow the rules and obey the standards." In fact the mixed performance record in meeting standards is part and parcel of the permanently incomplete design of the HRN, the non-fungibility of reliability in real time, and the need for market fallbacks in case of enduring high system volatility—all and more of which drive the HRN into real-time performance modes heavily reliant on informal networks, communication and coordination. To repeat, the best way for regulators and politicians to improve reliability standards and criteria is always to ask first: How does the proposed change reduce system volatility, increase network options, and/or enhance the focal organization's ability to shift and adapt in order to meet balance load and generation at any point in time and over time? We return to this point in Chapter 10 Recommendations.

Redefinition: What is Reliability in the HRN?

"Reliability consists of security and adequacy. Security is the ability of a system to withstand shocks and fluctuations. Adequacy is the ability of the system to meet its maximum load-demand. So, for security reasons, you need ancillary services," sums up one engineering expert. Other common measures of reliability mentioned by our interviewees include: risk (the probability of failure times the impact of failure), outages in terms of number and duration per unit of time, and regional or international benchmarks. Each measure, of course, has problems. As an electricity expert told us, "Traditionally we have looked at every failure we knew was possible. The problem is that there are hidden failures in the system, e.g., a protection device works one way when tested, but does not manifest certain modes until it is actually used in practice in interaction with other events."

In reality, the answer to "What is reliability?" depends on whom you are talking to and the criteria they are using, though more differences are found for service reliability than grid

reliability. Such differences can be found throughout the California electricity system, as our interviews show at length.

There are differences of view on reliability between the control rooms and their wraparounds in the HRN. A senior ISO scheduling official observed that, from his devil's advocate position, "nothing has changed" with respect to reliability, but that "shift people [in the control room] don't agree. They feel a lot has changed [in terms of what operators are asked to do]. Talk to engineering [in the wraparound] and they will say nothing has changed [with respect to grid reliability]." Within the ISO control room itself, differences exist between operators. We asked a shift manager about what kind of working relationship the manager had with the WSCC security coordinator located in the control room,

. . . he doesn't work for me, we coordinate. They can issue directives to us, he's oversight, if he sees that we do something that is not reliable. A 100% of the time we are already working on whatever the problem is when we receive the directive. It is his job to tell us this so as to make sure we do everything possible. Yesterday, generation dispatcher had difficulty in meeting the load; we met the load but were deficient in reserves...[The WSCC security coordinator] gave a directive to increase reserves. Does he see himself as part of the team? Yes, even though technically he is not, but physically he is. If he sees problems or something different, he will share info. He joins [our daily group control room meeting].

Since the ISO is the focal organization for reliability, it is important to ask, Just how do its control room operators define grid and service reliability?

"The ultimate goal here is reliability," says a senior engineer with ISO. "Nobody watches reliability but us," says an ISO market operations official. In ideal terms, reliability for ISO operators is equivalent to a good day in the control room, that is to say, equivalent to meeting peak load requirements safely throughout the day—and the longer the better—under "just-in-case" performance conditions. As we saw earlier, these days have become fewer and fewer since the crisis.

Control room operators over strongly agree over what qualifies as a good day. It is one where there is no shedding of load, no software glitches or breakdowns, load and resources are balanced, the ACE is controlled, frequency is kept at 60Hz, there are no path problems or violations, sufficient bids are in the BEEP stack and enough energy is on the ties. The truer test of reliability is how operators handle a bad day, and indeed bad days can end up good ones, when decisions are made which keep the system from failure. "In my mind we can have an absolutely horrific day out here that still end a good day. If we get through a 45,000MW day, if we didn't walk into many corners we couldn't get out of, that's a good day. In fact, three days that were the most horrific in my career turned out to be good days," according to a shift manager in ISO control room. Good days can of course finish badly, especially if operators end up with little or no room to maneuver. Understandably, a core element of real-time reliability is having a way out if something untoward happens, or, more formally, having flexibility through adaptive equifinality for the purpose of "just-in-time" performance. Because not all days are good ones or good ones in the same way, reliability is opened to being defined relative to the changing performance conditions ISO operators face.

A good day when it comes to reliability has formal elements. A recent ISO analysis of specific control room tasks tried to codify what reliability means in terms of determining day-ahead market ancillary services (A/S) as follows:

- Estimates of operating reserve requirements satisfy MORC and local area requirements
- Replacement reserves procured in day-ahead market are sufficient to satisfy system load predicted in the next day load forecast.
- Estimates of regulation reserves required in day-ahead market satisfy needs of system and local area reliability.
- Adjusted geographic dispersion of A/S is determined in accordance with good operating practices; and
- Amounts of A/S deferred to hour-ahead market ensure the most efficient use of available bids...

Somewhat less formally, a senior ISO engineer felt there are four components important for ISO service and grid reliability: First, “you have NERC and WSCC criteria and stability limits, generated through experience,” along with related measures. Second, “you have transmission grid reliability,” such that “when a piece goes out—one of the limits is breached—can you still provide reliability?” The “third part of reliability is proper maintenance, and the possible fourth is building and planning for future.” That “possible” sums up a great deal. A theme of this report is how real-time considerations have attenuated, if not actually driven out, the longer investment horizons for the grid and service reliability.

Few interviewees would gainsay that the private sector motivation to make profits puts system reliability at risk. “Market participants now have more profitable interests than reliability,” in the view of an energy expert on the public utility commission of another large state. As for tensions within California’s HRN including those instigated by private generators, one shift manager in ISO’s control room observed,

The markets definitely do have impact on reliability, and companies like ISO have to be very careful to structure their rules to minimize the negative effects of markets on reliability. In older system, I told generators what to do. In the new system it’s all based on the generator’s willingness to accept the price, and often times the results of that are not satisfactory for reliability. A system has to be very careful in crafting its rules, first to ensure that rules exist, for generators to follow the rules of the control area, and most importantly, the rules are enforceable. One specific example is a generator owner had generators online that were fast running out of NOx [nitrogen oxides] emission credits and we ordered the generator owner to close down and they refused to do it. We weren’t in a problem of losing the generation. Are we seeing more success in improving the enforceability of the rules? Not yet.

“Only the control area operator is on the hook for reliability. The merchant plant is only out for the profit motive,” a senior ISO engineer felt. We were able to interview a senior executive of an energy supplier who restated that position for us. From his perspective, reliability of the electricity generation is not as important as its commercial availability in terms of forgone opportunities and incurred penalties. Now, a megawatt is always available for its highest best use

measured in terms of price. That is, a megawatt contracted for tomorrow at X price would be better routed to an alternative purchaser willing to pay Y price at that time, if the latter price was higher than X and any penalties incurred as a result of not meeting the contracted X price. In fact, the older, pre-deregulation standard of reliability as high percentage availability of plant, whatever the demand, is contrary to commercial availability. The energy supplier was not primarily concerned with overall grid reliability, except to the extent that load swings and interruptions on the grid affect commercial operations, and even here there may be a positive spillover to the extent that reduced capacity of the grid to carry electricity may represent opportune times for plant outages and repair. "The only way the system can be at its best when peak periods roll around is when the power plants take advantage of the off-peak times by shutting down and repairing equipment that could potentially take them out in the peak," said the senior executive.

If such corporate views are representative, and we believe they are, then what has evaporated with the processes set into play by restructuring is any one single, gold standard of reliability. What has happened instead is the differentiation of reliability rather than the disappearance of reliability, save only in a unitary homogenous sense.

The problem, from the operators' perspective, is that privatization flies in the face of what they have learned it takes to make the grid and service reliable. A security coordinator phrased it baldly, "The single most frustrating thing for me about my job is the lack of control. In the old days, I could just tell the generators what to do and they would do it. This can't happen anymore. What's really wrong is that 40 years of experience, basically since the New York blackout of the early '60s, is now being ignored. It's been all for naught since deregulation." Another WSCC security coordinator went on, "Look at some of the white papers about this [deregulation] written by economists, who fore-fathered the concept. They expected everybody [generators] to be honest. Generators were expected to do what they were asked, and all that has proven to be not the case."

That said, where the specific control room operator stands on reliability depends on where he or she sits in terms of being responsible for it. A senior utility official gives the following account,

The ISO used to call them [utility's control room] up at 7.30 every morning and it was doom and gloom. [They] learned that it was the 10am call from the ISO that was the more realistic one. ISO would call up and say it was a no-touch day, i.e., a day when blackouts could occur. So [the utility] would send its maintenance people out to the switch centers to manually wait there to turn off stuff if needed. But the day would go by and no blackouts called, so [the utility] not only didn't get the maintenance done that these guys would have but also they didn't send the bill of the wasted hours to the ISO. What goes around comes around, that is, ISO would pass on these costs to the utilities which would then have to pass them onto the consumer. So what [the utility] did was run a program showing that they could shed the kind of load the ISO was talking about within 10 minutes through use of the [its automated load shedding] system. ISO accepted it, so the guys are no longer out their staffing their stations in no-touch days.

We, however, observed one instance where ISO operators did not want this utility to use its automated system during a Stage 2.

Redefinition: Where is the Reliability?

Reliability is in the grid and in the service, and since grid and service have been decoupled and altered under deregulation-induced performance conditions, so too has the nature and definition of the reliability.

Let's start with the grid, or more properly, the Electrical High Voltage System (EHVS), which with some exceptions goes down to the 69KV lines, where the radial (distribution) system begins. Not many people understand the importance of the latter with respect to reliability, as the DOE expert points out,

Technically, an individual's experiences with reliability problems are mostly the result of actions that affect the local distribution system, not generation or transmission. Tree branches falling on distribution lines, animals touching a line, that stuff. I would say the distribution level is responsible for something like 90% or more of the outages experienced by individual customers and the transmission for the remainder. (Of course, distribution system events tend to be very localized affecting service to a small number of customers, while transmission system events have the potential to affect service to a large number of customers.) Also, there are a huge number of electrical events that cause people to lose some of the services of electricity without electricity itself, i.e. without losing power. A change in power quality, what we call brownouts, can dim your lights but can really screw up your sensitive electronic equipment. These are typically the result of short term voltage sags. If you are a factory, losing power for a second can be as bad as having an outage. Power to your machines may be restored, but you have to first clean up the mess, and the restart your machines. The bulk transmission reliability problems that we study are ones that could lead the large scale outages. Most of what I'm talking about is bulk system reliability [the EHVS].

This EHVS grid imposes its own physical and technical constraints, limits and logic with respect to grid reliability. Source must equal sink, load equals generation. "One of the problems that happens is what we need is a balanced schedule, so that load meets need...If they don't get balanced, then we have to go into it and force the balance," says one ISO training official about the grid's logic. Also, when restructuring changed the grid, it redefined where the grid's weak-points are. In reply to a question about whether or not deregulation had changed the technical core of grid operations, an ISO transmission planner told us,

We have seen a dramatically different use of the system under the markets. People expected change, but not this much. . . Engineers always like to do new things, so that is also going on, e.g., innovations in static condensers. What we see more is that they are cutting all the slack out of the system, using RAS [automated responses to system conditions] much more. That is part of the technology fallout from market change. The weak points have changed. The generation dispatches have changed.

Consider what has happened to make Path 15 one very important weak-point. Once a way to assist the two separate grids of PG&E and SCE in cases of emergency, Path 15 has now become the shared organ of the Siamese twins borne out of the restructured ISO-managed grid. The crisis has made Path 15 the topic of cocktail conversation across America, as one of our interviewees put it. There are other ways to change the grid physically and thereby redefine reliability in the

process. Imagine if the distribution grid, which remember accounts for most of the outages customers experience in the course of their service, should become part of the wider reliability equation. Or imagine converting to a distributed generation system, where mini-grids would be managed by mini-ISOs, each with its own reliability criteria and standards. A University of California engineer and well-known observer of the California electricity sector sums up some of the engineering changes affecting grid (including transmission and plant) reliability this way,

The whole concept of reliability is changing. . . .The concept of security is changing. Security is ultimately still going to be guided by engineering standards and principles. They have a long experience to keep the system from collapsing. But you still have the problem of defining what exactly is a collapse. There's the issue of the quality of power that can be handled, its tolerance and frequency. People don't like the lights going dim. There is always the question if the quality can be moved to the customer side. Then you would have electronics installed in site rather than providing high-quality power throughout the grid all the time.

Not only has the grid changed as a result of restructuring, so too have the services reliably provided by that grid. In the words of a DOE energy expert, “deregulation has dramatically increased the volume and bulk of transactions” across the California grid. The increase has enabled greater differentiation in electricity reliability, be it for the Silicon Valley computer industry or your VCR. Moreover, as with any single resource that provides multiple services—be it large water supplies, telecommunications, the internet—the more services that the grid is expected to provide the more high reliability management of it will be required, e.g., in using transmission lines for internet services (Schiller, 2002).

A complicating factor important for reliability is, nevertheless, that the services provided by a reliable supply of electricity are not always clear-cut. Load is highly undifferentiated, and it is not evident just what services load represents nor for that matter whether these services are reliably “firm” or not. Restructuring has been much more effective in differentiating generation sources and where each is coming from and the risks associated with that generation than in differentiating the load is being served, apart from where that load is.

Again, Has Reliability Gone Down?

Since each criterion and standard of reliability has problems and the definition of reliability differs depending on where, when and who one is in the California electricity system, it is not surprising that there are differences over whether service and grid reliability changed as a result of restructuring and the crisis. “Is reliability better under deregulation? If it is measured in terms of not impacting on neighbors, yes it is even higher,” asserted a senior ISO official. “The ISO has never not met its monthly requirement of 90% violation-free operations. They were close but there is nothing wrong about that. Other control area operators are violating more,” reported a WSCC security coordinator. The senior ISO scheduling official asserted, “Reliability-wise we operate the system just as we did 10 years before. I piss off people here, saying that the reliability side remains the same. They say it has changed. Yes, they are now looking at the operating reserves to be changed. But that was coming without the crisis here, because there is no data behind it, the OR [operating reserve] criterion. I said, show me the technical data behind 7% [OR

criterion], how do you know it is right? A WSCC guy just said, because it has worked years. But how do you know it is the right number?” A senior grid operations engineer at PG&E told us, “Reliability criteria have actually become more stringent. Now you have more simultaneous double contingencies to have to be prepared for...”

Yet there is significant agreement that reliability standards are changing, and so too is there a sense in which insiders to the HRN and RRN feel old-time reliability has declined. When asked about what had happened to path violations during the crisis, an ISO shift manager reported, “Before November 2000 half dozen times, after November about two dozens times. Common denominator: lack of resources, the inability to dispatch resources to mitigate for flows on a path.” According to an expert energy engineer, the grid used to be reliable enough to back up the Diablo Canyon nuclear reactor, but with the electricity crisis worries had increased over whether the reactor should not have back generation uncoupled from the grid. “In the past, the assumption has been that grid reliability must be nearly perfect in order to keep nuclear power plants safe. They have emergency diesel generators as backups, but they were not as reliable as the grid. Now, with the blackouts, you have to wonder if the backups are more reliable. . . The concern is that grid reliability could cause nuclear power plant accidents and that the result would be catastrophic,” he said.

Certainly, the political view is that reliability declined as a result of restructuring and the crisis. A senior legislative aide asked, “Has reliability gone down? Yeah, it would be hard for anyone to conclude otherwise, especially with rotating outages in the last months.” “Our transmission system is third-world right now,” in the view of a senior regulatory staffer. In the view of a senior ISO engineer, “I’m certainly less convinced than when I started [here at the ISO] that reliability through markets works.”

The preceding chapters and Appendix A of our report provide evidence that service and grid reliability, in terms of shedding load and controlled blackouts, was much higher than perceived or theoretically thought possible during the electricity crisis. What should be noted here, however, is a subtler point. Namely, one is much more certain in arguing that the value of reliability went up rather than that the level of reliability went down as a result of restructuring. One reason why the generation plants and stranded assets were so valuable, once the restructuring process had begun (Appendix A), was that they were backed up with a *de facto* guarantee that government would ensure reliability in the electricity sector when push came to shove, as it indeed did during the California crisis. Similar premiums have been observed in other deregulations as well (Martin, 2001). Restructuring generated new performance conditions that tested the value society places on service reliability and grid reliability, even when the former poses threats to the latter.

Perhaps the best way to judge the extent of the redefinition of reliability as a result of the restructuring and its induced crisis is to compare and contrast the principal features of the older HROs and newer HRN for achieving this same end, balanced load and generation (Table 8-1).

As can be seen, reliability of grid and service entails some very different features for HRNs than it does for the predecessor HROs, such as Diablo Canyon nuclear generator and some of the earlier integrated utilities. While some features remain the same or similar, there are appreciable

differences with respect to performance, culture, non-fungibility, trial and error learning, and redundancy, at least when the core reliability requirement is balancing load and generation reliably. The reader sees the quick hand of real time in this rewriting, points developed more fully in Chapter 7 on the many factors pushing and pulling the system operator into real-time reliability as a result of restructuring and crisis.

Table 8-1
Comparison of Selected HRO and HRN Features

HRO Features	HRN Features
<ul style="list-style-type: none">• high technical competence	<ul style="list-style-type: none">• high technical competence
<ul style="list-style-type: none">• high performance and close oversight	<ul style="list-style-type: none">• at times mixed performance and little or no RRN oversight
<ul style="list-style-type: none">• constant search for improvement	<ul style="list-style-type: none">• constant short-term search for improvement
<ul style="list-style-type: none">• often hazard-driven adaptation to ensure safety	<ul style="list-style-type: none">• often hazard-driven adaptation to ensure safety
<ul style="list-style-type: none">• often highly complex activities	<ul style="list-style-type: none">• often highly focused complex activities
<ul style="list-style-type: none">• high pressures, incentives & shared expectations for reliability• culture of reliability	<ul style="list-style-type: none">• shared professional norms, competence, information and accountability substituting for the culture of reliability
<ul style="list-style-type: none">• reliability is non-fungible	<ul style="list-style-type: none">• reliability is non-fungible in real time, except when service reliability jeopardizes grid reliability
<ul style="list-style-type: none">• limitations on trial & error learning• operation within anticipatory analysis	<ul style="list-style-type: none">• real-time operations necessitating improvisation and experimentation• operations outside analysis
<ul style="list-style-type: none">• flexible authority patterns within HRO	<ul style="list-style-type: none">• flexible authority patterns within focal organization and across HRN
<ul style="list-style-type: none">• positive, design-based redundancy to ensure stability of inputs and outputs	<ul style="list-style-type: none">• maximum equifinality (positive redundancy), adaptive equifinality (not necessarily designed), and zero equifinality, and depending on network performance conditions

Implications

Does this mean that reliability has lost a good part of what had been its enduring meaning as a result of restructuring? Late one evening in the control room, the coordinator of the hour-ahead market desk was describing the kinds of problems they faced in keeping control of the ACE: CPS2 violations, computer failures, software disasters, path problems and violations, data problems, late submissions by SCs, not enough bids in the beep stack for incing or decing, ignoring dispatch orders, shedding load. . . The list of problems kept expanding to the point that one of us typed into the interview transcript: “How the hell does this add up to reliability?!”

What has actually happened is that changes in system volatility and the variety of available network options to respond to that volatility have institutionalized and sanctioned— notwithstanding formal procedures, standards and criteria to the contrary— “just-in-time” performance, “just-for-now” performance and “just-this-way” load shedding as “legitimate” ways of meeting the reliability requirement, even though these performance modes depend

heavily on dangerous-to-codify informal networks, relational communications, and experiential/institutional knowledge bases.

Furthermore, the drive to real-time reliability demonstrates how crises or the potential for volatility-induced crises have become unexceptional. "Just-in-time" performance is not an odd, one-off, unusual and irregular event any longer. During our stay in the ISO control room we observed the quickened pace of activities and many other features of "just-in-time" performance in multiple, repeated versions: during an hour's ramp; in the early morning and evening when electricity usage increases considerably; in the hour ending 23 when load is going down and generation is going up during the ramp; during the weekdays as distinct from the weekends; during summer months in comparison to the rest of the year; during software upgrades and glitches in the control room that posed potential interruptions in service and/or grid reliability; and during scheduled outages that took place throughout the day, week or over months. A year after our control room observations, operators were still talking about "just-in-time" and "just-for-now" performance as their "normal" modes for balancing load and generation.

The search for flexibility and adaptability that characterizes "just-in-time" performance is something that is going on in some way, shape or form during the day, most days. In the past, it took massive blackouts, like that of New York in the 1960s, to change reliability standards. Restructuring represented another punctuated equilibrium, this time with often involuntary learning over just how far operators can push the technology, markets and their fallbacks without losing control of the ACE and paths in the process.

Second, the earlier focus of HROs on fixed standards, limits and bandwidths within which to make adjustments for stabilizing inputs and outputs obscures the fact that factors, such as operating reserves, have always been dynamic rather than static. Reserves are continually moving, very much a function of the single largest generator online, or the intertie with the most imports, or the actual load during the day. Add the fact that generators, lines and equipment all have their official versus actual operating limits dependent on a host of very local factors, all with their own standard deviations, and the dynamic nature of the management moves from background to foreground. To choose randomly from one of the many descriptions given to us, the "Path 15 real-time operating limit is a 'moving' limit," according to the recent ISO task analysis, "that depends on the following current operating conditions: actual West of Borah [WOB] flows, level of pumps/generation/load to be dropped by the I RAS, ambient temperature at Gates substation, the '300 Criterion' (which allows the limit to increase for low WOB flows and other requirements), and equipment out of service, which derates the operating limit." You get the picture.

Indeed, such site, time and scale-specific factors explain why we were often told that the balance of load and generation must be managed dynamically. It is not static and never has been; management has always been "primed" for change, including those in the criteria meant to govern that management. What is new with the advent of restructuring is that the dynamic nature of the management challenge has qualitatively and quantitatively changed in real time. "The difficulty is that you now can see flow patterns [on the grid] that you have never studied, and so, based on traditional indicators, you may not be able to tell whether you have a problem," according to the DOE expert. "Nobody saw 3000MW ramps before [deregulation], line flows are

different now, Diablo [Canyon reactor] power goes south to Los Angeles, no longer to San Francisco. Things flow and act different,” in the words of a senior ISO engineer.

Also specific changes in the RTE and RRN redefine reliability for the ISO. First, conservation in the RTE has itself become part of the HRN reliability equation. Instead of expecting the HRN to provide reliable services to us, we in the RTE are expected to make those services reliable in important ways, through conservation and, in the future, through residential metering. Customers are now told that if they want adequacy and security—the engineer’s version of reliability—or services during peak load times—the market’s version—then they must become part of the HRN’s reliability equation for balancing load and generation.

Second, the ISO is the manager of the grid, not its owner. If it is important, as was so often repeated in our interviews, that generation is no longer owned by the utilities, then in parallel fashion the fact that the utilities own, but no longer manage their grids must also be significant. As we know from the economics literature, the divide between owners and managers raises a host of issues at the interface between public and private sector interactions. In the California case, our interviewees agreed that what has fallen between the cracks of this owner-manager divide is maintenance and longer-term investments in the grid. Consequently, we can expect grid reliability, and thus ultimately service reliability, to be challenged in the future, as the lack of investment poses the double whammy of increasing volatility and decreasing options variety for the ISO.

Third, the grid and the technologies (including market software) which the ISO manages are continually changing and redefining reliability. “Ultimately, we are looking at more intelligent, automated and decentralized tools to help control the reliability of the grid,” said a DOE expert. Change the locus of where the balance is made and you change the nature of the reliability achieved. Fourth, it should go without saying that the use of this as well as other software providing “real-time” information is likely to reinforce real-time reliability approaches, not least of which is “just-in-time” performance. The conflict between RRN and ISO standards for reliability is especially visceral when the issue is real-time reliability. The WSCC criteria tell us to stay within prescribed bandwidths, a senior ISO official told us, “But we ask ourselves at what risk if we did meet these criteria and caused lights at intersections to go off?”

Finally, we need to say something about luck and its relationship to reliability. Ensuring reliability under real-time performance constraints, as operators will tell you, has always involved an element of luck, and deregulation has made that more true than ever before. Why? Because the risk of failure is now much higher than before restructuring. From the control room operator’s perspective, good luck is the non-occurrence of failure in the absence of exercising failure avoidance options, while bad luck is the occurrence of failure in the presence of exercising failure avoidance options. Restructuring not only increases the number of different ways to fail but also ambiguity over what options are really available for failure avoidance. Thus, the operator now faces a performance conditions where good and bad luck are more important than ever before.

What to do? At least two factors are involved. First, in one sense, luck—whether good or bad—is really just another way to experiment in the face of real-time pressures. A senior grid operations engineer at PG&E described the following near miss,

. . . it's a hot day in May 1987 and we had three pumps out. We nearly caused a voltage collapse all over the western grid. Everything was going up and down, we were trying to get power from all the nuclear units in the western grid. Life flashed before our eyes. And then the gen dispatcher did intuitively the right thing. He said, Shut one pump down. How he figured that, I still don't understand. It was something we had never seen before. We had no procedures. . . We went back and looked at it, and the planner said, Oh yeah, you should never have been running three pumps, and we said, Where did you get that from? So we started writing new procedures.

What operators call luck, in other words, is real-time improvisation, though understandably involuntary and on a larger scale than most HROs (and even HRNs) prefer.

Second, luck is the other face of real-time confusion and incomprehension when messes like that faced by the operator go good when they could have gone bad. In recounting one bad day that turned out good, a shift manager in ISO's control room described how "just by sheer stroke of luck I had made a voltage change at that time and caught what happened." Luck here is a way out for operators when they are in a bad corner, an equifinality created by the multiple, concatenating variables of just-in-time performance that afford chances to turn bad messes into good ones.

Having reliability criteria flexible enough to enable operators and their wraparound to take advantage of the lucky opportunities that will inevitably albeit never reliably come their way will be crucial to future grid and service reliability under restructuring. If luck is part of the equifinality of options, and adaptive equifinality is important for just-in-time performance, then control room operators and their wraparounds will have to make their own luck during the highly volatile times ahead.

9

THE PUSH AND PULL TO REAL TIME

Introduction

The preceding chapters have had a good deal to say about the factors and processes underlying real-time reliability, that is, "just-in-time" and "just-for-now" performance in order to balance load and generation under HRN conditions of high system volatility. This chapter, the longest in our report, summarizes and elaborates that material with respect to sixteen factors pushing and pulling system operators in real-time performance conditions. The notion of "real time" is somewhat elastic, but for many operators it means at least focusing on planning and managing operations for the hour ahead or within the current hour. While the chapter covers some of the same ground as the preceding material, it introduces new points and provides a comprehensive overview of the multiple reasons why real-time reliability conditions are here to stay in California for the foreseeable future.

Push to Real Time

One storyline in California's electricity restructuring and crisis is how a large part of the electricity provision ended up in real time. As we have already seen, market design unintentionally provided many incentives for generators and the distribution utilities to defer energy transactions until the last minute. With supplies becoming scarcer, our patient, the market, was wheeled into intensive care, as strong pressures of the generators to withhold to the ISO's real-time market increased dramatically. At the start of the crisis, they could get higher prices as well as capacity payments. The capacity payments meant they then got paid twice: first for providing the generation as capacity, that is, to keep them available for ancillary services. Then when the generation is actually used, they would also be paid for the actual generation. (The double payment has been mitigated since to a FERC ruling in April 2001.)

Somewhat less predictable were the incentives for distribution utilities to move to real time. First, there is the much-criticized CPUC position that dissuaded utilities from seeking to obtain regulatory approval for forward contracting. So their load had to be covered through the PX. The market clearing price model used in the PX meant that, even when prices were higher in the real-time market, it may still be rational to move part of the load there as part of a profitable arbitrage strategy. A very different incentive was the fact that in the real-time market, the ISO bought the electricity on behalf of the utility. While in the end the utilities also pay for these transactions, they are not subject to the same CPUC purview as the transactions the utilities undertook in the forward markets. In buying and managing electricity the ISO is not under CPUC scrutiny, but rather regulated by FERC.

There were more incentives at play, but for the ISO the net effect was certain: They were pushed to work in real time and exposed to significantly higher system volatility. A senior ISO control room official compared this present with the pre-restructured past,

To operate the system now, it's so dynamic. It's a full-time challenge. It's very volatile. . . The biggest change are all the unknowns in all your decision factors and forecasts. A lot of decisions have to be made fast, in a short time. I have no guarantees that my forward schedules are right. I have no idea how much imports I can use for balance. There is a lot of intuition going on, and a lot of experience coming in.

It was more stable [in PG&E before deregulation]. At PG&E you had the generators under control. You knew the day before about today's capacity. Forecasting tools [were] better because you were dealing with a smaller grid and thus more precise. You knew supply and long term deals that I don't have know. Some of us thought it was important to maintain long term arrangements but the marketers wouldn't have any of that. We used to have arrangements with BPA. We would give them surplus in the fall and winter and get 1,000MW during peak during the summer. Now we can't do it. It makes our job a lot tougher. It's the biggest challenge in my career.

As discussed in our report's framework, real time refers to planning for the hour ahead or within the current hour for operations. All four performance modes—"just-in-case," "just-in-time," "just-for-now" and "just-this-way"—therefore have elements of real-time operations in them, albeit the importance of real-time operations vary dramatically. For our purposes, the push to real time is best thought of as the shift from "just-in-case" performance to other performance conditions, especially those of "just-in-time" and "just-for-now." "Every day you just have to wait and see what the situation is," said an ISO staff member familiar with its control room. For two ISO operators, "every day is so different." So too for a PG&E control room operator, "Every day something is different. It's like facing 365 different pitchers. Everybody's got a fast ball, but it's their own."

For the purposes of this chapter and report, real time is primarily characterized through the left-hand side of the Figure 2-3 framework where system volatility is its highest. Something unpredictable or uncontrollable happens, and situations of low volatility and high options give way to increased volatility and declining options, where the time for planning is foreshortened, the pace quickens and the balance of load and generation is confirmed hour by hour, if not up until "the last moment." Explaining why it was ludicrous to give twelve-hour notice of blackouts, an ISO shift manager said, "That's just not possible. It's the same as telling them what load I'm dropping next year and notifying them [now]. These people don't understand that you can't just order a truckload of megawatts. This is real time!"

In such a setting, the longer-term horizons evaporate as operators are pushed into the urgency of operating hour-by-hour, minute-by-minute. During the crisis, the operations of the day-ahead and hour-ahead markets took on less urgency as the operations of the BEEPer, gen dispatcher and schedulers for the interties become much more pressing and compelling. "Real time always takes priority," as a senior ISO engineer put it.

The fact that a large portion of overall electricity was provided at the last minute during the crisis has been read as a decline in service reliability by many insiders we interviewed, who felt pushed into real time as the last resort to prevent load shedding or worse yet, islanding, in order to ensure grid reliability. What this leaves unanswered, however, is how the ISO and with it the HRN managed to be as reliable as it did, given that so much depended on real-time operations. To understand that, we have to see that there are also pull factors to real time, even for the ISO.

Pull to Real Time

Let there be no misunderstanding: If the ISO had anything to say about it, it would never choose to be so exposed to the volatile conditions of real time. Formally, the only real reason they operate there is because there were pushed into doing so. "I learned about Murphy's law," reported a senior grid operations engineer in PG&E, "If the worst can happen, it will happen. Which makes sense, because the worst is most likely to happen when the system is at its most stressed. Reserves are needed just when maximum transfers are happening." A recent National Research Council report (2002, 6-4) explains, "A highly stressed system (e.g., if power imports are high and transmission reserve capacity is low. . .) would be more vulnerable to cascading failures and the resulting longer-term blackout."

But there are also pull factors at work and to understand them we must first start with the fact that the ISO has developed performance strategies to survive under "just-in-time" and "just-for-now" performance conditions. These strategies, as described below, cope with real-time pressures to balance load and generation. In fact, as we see, these strategies work better than some more conventional ones do when system volatility is lower. That is, when exposed to high volatility, such as prices paid on the spot market, the existence of strategies whose very success depends on their being real-time conditions in place serve as pull factors to "just-in-time" and "just-for-now" performance. Becoming a successful professional in crisis management has its own attractions and that, too, is part of the story about how the system ended up and survived remarkably well in real time.

While Governor Davis may believe that he is the reason why the lights stayed on during the crisis, the following sixteen factors actually account for why grid and service reliability were as high as they were during the California electricity crisis. The factors are set out in Table 9-1 and discussed in the following subsections.

To repeat, the point we leave the reader with is that control room operators in the HRN (not just its focal organization) are not only being pushed into to real time. They are also being pulled for quite positive reasons, as it is in real time that operators can balance load and generation using strategies they could not use under other performance conditions. Thus proposals that are geared to lessen volatility, such as new generation coming online, have to be assessed in terms of whether they also increase ISO options in such a way as obviate the need for and appeal of these other real-time strategies for the control room operators.

Table 9-1
Principal Push and Pull Factors to Real-Time Reliability

- Real time is an answer to persistent network incompleteness
- Positive interdependencies are most visible in real time
- Real time remains informal, non-routine and relational
- The need for small output variance (load and generation must be balanced) is adapted to by real-time multiple means with large process variance (equifinality)
- Real time legitimates and accommodates redefining reliability criteria
- Reliability is non-fungible in real time
- Real-time operations justify improvisation and experimentation
- HRN tolerance of what would otherwise be considered low performance is greatest in real time
- While often negative, strategic behavior can have positive effects on real-time reliability
- Markets and deregulation incentives exist for ensuring real-time reliability
- Pulling new participants into the HRN becomes an option for increasing real-time reliability
- Shared professional norms, competence, information and accountability substitute for the culture of reliability in real time
- Reliance on real-time information and software leads to reliability in real time
- Flexible authority patterns and teams become the locus of reliable behavior
- HRN relations with the RRN and RTE seek to foster and protect real-time reliability
- Unbundled generation can increase predictability, transparency and self-organization in real time

1. Real Time is an Answer to Persisting Network Incompleteness

As described in Chapter 5, when a network is created, it is never fully designed. If all designs are incomplete and that is a constant, then considerable weight ends up being put on real-time activities to ensure the reliable provision of services. In real time, almost everything that matters is denominated in terms of load and resources, while performance standards are their clearest when it comes to frequency, ACE, CPS and path violations. There is incredible pressure to pull the HRN together at the last minute, precisely when success or failure is most visible and salient to everyone concerned. In the words of one of the ISO's lead officials in marketing operations, "in real time everything comes together here in the control room."

Design incompleteness is thus a double whammy for the control room operator: It increases volatility while leaving options variety in question, thus encouraging the operator to put off

decisions to the last moment when information has the greatest probability of being most reliable. “Real-time price is the benchmark for knowing what we do,” said one of our interviewees.

What that means is that all electricity systems, as long as they face technological innovations, new markets and changing reliability standards, will remain incomplete and thus a significant part of high reliability will be in real time, whether the HRN, RRN or RTE likes it or not.

2. Positive Interdependencies are Most Visible in Real Time

None of the other control areas in the western grid can really afford the California grid being unreliable. Most immediately, there is the grid dimension of networked reliability: Instability in the California grid (including plant and transmission lines) can ramify and cascade into instability throughout the western grid.

In this kind of setting, positive interdependencies come to the fore when they are needed most. A senior ISO control room official pointed out that in the one month during the crisis BPA bailed the ISO numerous times at the last minute, because BPA knew that California would need to go the extra mile later for them later. In his words, “the rest [of the western grid] can’t afford California to go down. BPA knows that it needs us in the fall.” There are times when different control areas are chasing the same megawatt, but positive complementarities between control areas are their most palpable and evident in real time.

But there are no guarantees here. A year after our above interview with the senior control room official, he and others informed us that it was increasingly difficult to depend on adjacent control areas to bail them out at the last minute with imports.

3. Real Time Remains Informal, Non-Routine and Relational

When incompleteness in design reinforces the importance of getting it right in real time and if and when strategic considerations across the western grid reinforce real-time reliability in the California portion of that grid, many of the operational activities in real time have understandably remained informal, tacit and unofficial.

We have spent considerable time in this report, particularly on the topics of wraparound and always-on management, underscoring the importance to high reliability of flexible authority patterns that revolve around operators communicating and working in concert with each other in the face of real-time performance constraints. To summarize, real time values and privileges the non-routine over the routine, the informal over the formal, and the relational over the representational. Much is handled as a last resort, hour-by-hour and in the last minute, under volatile, contingent conditions. Operator behavior becomes necessarily provisional is best summarized in the phrase, “operators using their discretion.” A senior manager in the ISO operations engineering unit, responsible for a large body of procedures, told us “part of the experience is to know when not to follow procedures...there are bad days when a procedure doesn’t cover it, and then you have to use your wits.” Smart people like to use their discretion,

and it takes very smart people to run the control room in order to produce grid and service reliability in real time.

The mix of informal, tacit, non-routine and relational is best seen in the shifts to "just-in-time" and "just-for-now" performance under high system volatility. During "just-in-case" performance, the shift crew remains relatively static in the control room, as described earlier. One or two members in the wraparound may be activated to assist the BEEPer as pressures increase. But once that tempo and pace quicken under increased volatility, sometimes to the point where staged emergencies are declared, the control room team widens considerably to include an emergency response team of experienced engineers and higher level officials who have working relationships with large importers of electricity into California. The HRN itself takes on lateral matrix features, where staff across units in different organizations work as teams in the emergency. The capacity for key players to make informal mutual adjustments within the control room, its wraparound and the HRN is a core part of ensuring real-time reliability.

The combination of experience and institutional knowledge for operating within informal, non-routine and relational-demanding situations ("keeping the bubble") is particularly important for "just-in-time" performance. Real-time reliability is best seen as the result both of a high sense of urgency to be reliable and of the options to continue to be reliable, even if something fails along the way. In this way, the ability to adapt to the unexpected that comes from working familiarity, experience and knowledge with the grid and people operating it decreases the operator's dependence upon a single knowledge base or set of relations, procedures, or skills to ensure reliability. "Just-in-time" performance is having multiple means to achieve the single end of reliable electricity.

One way to sum up the informal, non-routine and relational is to say that much of real time for the ISO operator (and for others in the HRN) is manual: not by the manual, but manual as in hands-on. It is what they have to do when automatic market and technology elements fail or mislead. They go in manually, as when they call the generator directly or adjust the AGC. When asked what a bad day was, a senior ISO market operations manager said, "It would be when the software systems are not operating properly, and the control room operators must make manual dispatch decisions rather than using computer programs."

Others go in manually and force balanced schedules. Others make due with their wits, calculators and phones. For operators who place a great deal of value on control—it is after all, the *control* room—real time is when manual intervention and operations are most likely needed and useful. "When the plants are to be turned on, it can take about 8 hours to get them running. Real-time control is therefore critical and we need to let engineers do that until the market is much more developed," argued one of the experts we interviewed. In this way, RAS's stand out precisely because they are automatic and real time, without the option of going manual. "If something goes wrong with Path 15," explains a senior ISO official, "then that would feedback all the way to Borah Idaho through a loop and burn everything up there. That is why we have RAS in place, to respond very quickly to these things."

Against this background of a control room that runs on high-octane informal, non-routine and relational in order to maintain real-time reliability, the complaint against the ISO that it has

added another layer of administration and bureaucracy must be reconsidered. When it is recognized that ISO control operators came overwhelmingly from the HROs who no longer could be reliable in the way they once were because of restructuring, this “layer” looks also to be a source of real-time adaptation and on-the-spot management that our framework calls adaptive equifinality. This leads us to our next factor.

4. The Need for Small Output Variance (Load and Generation Must Be Balanced) is Adapted to by Real-Time Multiple Means with Large Process Variance (Equifinality)

The operational requirements of the ISO are so intense that they are unicentric—they are focused on overarching reliability concerns to such a degree that everything else is “politics.” In a term introduced earlier, the reliability requirement to balance load and generation is “always on” in the focal organization and much of the HRN.

But while the output of the ISO is clear—reliable electricity—the options and means to achieve that objective are multiple. There is large variance on the process side to achieve an output with little variance. That is to say, there are many more interventions, transactions and contingencies in the control room under restructuring than in the older utilities. A sense of this shift to wide process variance around a persisting low output variance is captured in the comments of a senior ISO scheduling manager,

The old way was nicer and easier, you had 20 years of experience behind you, but now you still get the [same] results. The fact that a lot of it has gone into real time has caused headaches, but that’s it. There was more tunnel vision before, scheduling was a narrower task. It has widened out, it encompasses more, pushing things into real time, there was never before a BEEP person, there wasn’t an hour ahead, the tunnel is widening a bit more, but concepts remain the same. . .

This large variance can work against achieving reliability—many more things can go wrong now—and there will be strong pressure for the ISO, from within and without, to reduce the number of variables with which it has to deal. Yet the large-variance/single-focus of the ISO also poses opportunities for real-time reliability in at least five ways.

First and most obviously, the large variance represents equifinality in achieving reliability under “just-in-time” conditions. There are more ways than one to ensure highly reliable provision of electricity, even at the last minute. As one ISO control room shift manager put it, “In this [control room] situation, there are more variables and more chances to come up with solutions.” “It’s so dynamic,” said one of the market grid resource coordinators, “and there are so many possibilities in bid structures. Things are always changing.” As we have seen, when the ISO cannot obtain the electricity through markets, it goes out-of-market. When it had a voltage problem in Fresno threatening reliability, for instance, the option was to take a line out or call up an RMR unit.

Second, errors and problems of varying magnitudes are visible in the control room since they related directly to load and resources, not as proxies for the system as whole. When the ACE

drops 2,000MW, it has to be a scheduling input error; when it drops 600MW, it has to be a generator, and very likely to be among a specific set of generators. While there are many ways in which load and resources can go out of sync (the large variance), the ways they go out of sync are clearer, more visible and, most important, often correctable only with more resources, less load or both.

The unicentric focus on load and resources justified by being real time means not only that the performance standards are their clearest then but that *feedback on whether these standards are being met is almost immediate in terms of frequency, ACE and CPS2 violations*. That is, feedback is clearest about whether the service is “on” or “off.” Operators in the control room can look for and read signature events around load and resources, while admitting they do not fully understand or comprehend what is happening in the entire system at that moment. In fact, “just-in-time” performance would not be “just-in-time,” if there were not some confusion about what is really going on in the network and wider electricity system. In the words of a WSCC security coordinator, “99% of the time things don’t happen in real time as scheduled.” “Generators are rarely where they are supposed to be. That’s the nature of the beast,” said an ISO engineering manager.

Feedback and pattern recognition when it comes to load and resources are more important to achieving real-time reliability than is (incomplete) causal comprehension of the system as a whole. The focus on load and resources, rather than the whole system, is a primary difference between the HRN and the older HROs. The HROs managed reliability as a base because it was the whole system that was in jeopardy from potentially cascading error. The newer HRN manages reliability at the margins of operating reserves and path limits, because that is where balancing load and resources is most dynamic in real time. The shift is possible because balancing load and resources has been a reliable way of avoiding cascading error. Of course, the related lack of maintenance and increased wear-and-tear on generation plant are growing problems. No one doubts that maintenance and investment in the transmission grid are truly The Cost Crisis of the Future, but then again no one in the control rooms doubts that it is the responsibility of others to worry about that. “Reliability is real time, but the generation and transmission systems you rely on take years to build,” in the words of the DOE electricity expert.

Third, large variance in terms of control operators experience is a resource when the unexpected and contingent are “normal.” After describing the wide background of his crew, an ISO shift manager summed up by saying that here was 200 years of experience here in the control room. Another control room operator in the ISO told us that the wide experience meant that it was much more likely that someone in the control room knew how to handle or deal with something surprising. Wide experience, combined with cognitive, institutional and experiential knowledge of the control operations by operators and the wraparound (achieved through training on-the-job and over multiple positions), have served to promote and reinforce reliability, particularly when the sense of performance urgency is high.

Fourth and elatedly, large process variance comes in the form of experience in the wraparound as much as in the control room. As we reported, some operators in the control room said they scarcely noticed the demise of the PX, the day after it went bankrupt. A variety of people in the wraparound (including those in market and grid operations), however, scrambled to ensure new

rules and procedures were in place to make the transition as seamless as possible. The wraparound works, when the variety of experience there buffers as well as supports the control room from an increase in real-time volatility due to exogenous events. Consequently, much of the wraparound takes on real-time concerns in its core activities, be it outage coordination and prescheduling on the one end or settlements on the other.

“Most all of [outage coordination] today is less than 30 days,” reported the senior ISO outage coordination official. Is that like the push to real time we see in other ISO units, we asked? “It is exactly like the control room in being pushed to real time,” he replied. Prescheduling too has become more last minute, while pressure remains to keep settlements within a ten minute range of actual purchase. “Preschedule is day ahead. Changes into the preschedule spreadsheet are real time. . . The ironing out is in real time,” a manager of real-time scheduling in the ISO reported to us. “The preschedulers leave tonight with 24 figures for each hour ahead, but when you look at what it is after the fact, the numbers are very different,” said a senior ISO scheduling manager. As for settlements, “When settlements are not processed, they notice that all the way back into real time, to the bids in the BEEP stack,” said a senior ISO official.

Fifth, part of the large process variance reflects the fact that the risks the ISO face are more differentiated than in the older integrated utilities. Speaking of private generators, a senior ISO outage coordination official said, “We want them to manage their risks appropriately so we don’t have to manage the uncontrollable risks.” Some risks are more manageable today or managed differently in comparison to the older utilities when it comes to achieving reliability. For example, since the grid is now carrying an order of magnitude more transactions of varying amounts than before restructuring, new weak-points, namely Path 15, have become more visible and therefore more the focus of remedy.

5. Real Time Legitimizes and Accommodates Redefining Reliability Criteria

The report has described the multiple ways in which reliability and its standards are being redefined. How such redefinitions reinforce the real time provision of reliable electricity is discussed below.

The emerging reliability criteria and standards have one common denominator: They reflect the effort by members of the HRN, particularly the ISO, to adapt reliability criteria to meet circumstances that they can actually manage, where those circumstances are increasingly real time in their urgency. What cannot be controlled "just-in-case" has to be managed "just-in-time;" if that does not work, performance has to be "just-for-now;" or, barring that, "just-this-way" by shedding load directly. In each instance, reliability becomes that bar that the operators can jump.

Thus, there is a paradox between having reliability standards and having multiple ways to produce electricity reliably. On the one hand, the standards are operationalized (in terms of the WSCC) and the fact that performance can be empirically gauged against these operational measures is the chief measuring stick of whether electricity is being provided reliably or not. On the other hand, the standards are everywhere being redefined, when not questioned, because there are an increasing number of times when only by pushing the standards to their limits and sometimes beyond are the lights kept on.

The combination of clear standards and their questionable assumptions serves to reinforce the importance of real-time provision of reliable electricity. For now it is not only the network that is incomplete but the standards as well (i.e., they are being redefined as you are reading this), such that it is left to control room operators to connect the dots at the last moment, when load has to equal generation. As you read before, keeping within the WSCC bandwidths is not necessarily the best way to save lives at traffic intersections. To put it differently, one pull factor to real time is the fact that real-time learning and experience legitimate the need for emerging new standards.

Thus, just as we can expect the network to remain incompletely designed for the foreseeable future, so too can we expect the modification of standards to be ongoing—and with both, the pressure to reconcile things hour by hour when the reliability requirement of balancing load and generation is its most compelling and unambiguous. Or to put it in HRO terminology, the closer you get to real time, the more reliability is treated as non-fungible.

6. Reliability is Non-Fungible in Real Time

To put it succinctly, operators are pulled, not pushed, to real time by the fact that in real time load has to be treated as all but given, where it is only resources (particularly generation) that have to be managed in order to meet the reliability requirement. Because options are limited to one side of the balance, the limits to choice become clearer, and clarity is important for just-in-time performance under volatile situations. In the very short term that real time occupies load is always there, always unavoidable, always to be matched with resources that can be managed at the margins. In real time, there really are no substitutes for the reliability of electricity as a service save one single exception: Service reliability is given up when it directly jeopardizes grid reliability. Service reliability and grid reliability are, in other words, their most non-fungible in real time. This is an attractive position for operators to be in, particularly when the high system volatility they are operating under is none of their creation but getting sufficient clarity to make a decision is all of their job. To put it more formally, real time is about exhausting all available options in the name of the non-fungibility of grid reliability—whatever else happens operators cannot let the grid island. Real time is about depleting positive redundancy and equifinality as a way of avoiding collapse of the grid.

Given long enough, of course, all fixed costs become variable. In the long term, fixed inputs give way to substitutes. No input is really non-fungible over the long haul, and there is no reason to believe the electricity sector is any different, given the dramatic technological innovations in that sector during the last century. Who knows, in the long run of political science fiction, we might not even need electricity as a society.

But real time is not the long-run. Indeed, if we had to define “real time,” it would be when reliability has ceased to be an input into electricity production and has become an output of the service produced, for which there are few, if any, substitutes in the very short-run in California. We all know by this point that electricity is not a quality of a service that can be substituted by other service attributes, when margins between load and generation are their tightest, as they are in “just-in-time,” “just-for-now” and “just-this-way” performance.

Those margins, moreover, are increasingly tight. “Every thing we do is so close, so close,” an ISO shift manager observed when they had just missed a Path 15 violation by seconds. “Path violations are the worse,” reports another shift manager, “If we have a path violation for 10 minutes and lose a generator or other large piece of equipment, we might slow down the whole WSCC [western grid] system.” A third shift manager describes the close calls and the effect of changing performance conditions on reliability as follows,

In my world, lack of generation is most important. . . . What did the last few bad days look like? We started close to the forecasts. Knew it was going to be close. Then we lost a 550MW unit that morning. We can deal with that. Then BPA cut us, without warning. Only a 20 minutes heads-up. Next hour WAPA, who was selling us energy, cuts 800 MW from that. . . We move from warning to Stage 2 and Stage 3 to send people out to get ready to shed firm load. How do you get out of that mess? How do you solve the unpredictability? We don't. I would take my own generation over imports anytime. It used to be used just to supplement in-state generation. Now we are heavily reliant. Now we have 4600MW import. If that is gone we're screwed.

We are screwed, because it is at such times that electricity cannot be substituted with or traded-off against electricity's other features, such as whether or not it is green energy. Grid reliability cannot be cashed out and replaced by dollars in the very short run that the ISO shift manager is describing. We sometimes forget that, since economics is a theory of service and quality substitutability, it cannot be a theory of non-fungible reliability. As said earlier, economic theory will have to be recast in order to show how the non-fungibility of real-time service and grid reliability is not itself a fundamental limit on market behavior in the electricity sector.

7. Real-Time Operations Justify Improvisation and Experimentation

The earlier literature on HROs found that they avoided anything like the large-scale experimentation we found in the California HRN. Yet the preceding chapters should leave the reader with little doubt that real-time “operations” in the ISO means literally improvisation and at times widespread experimentation. In terms of our framework, the improvisational and experimental are what we have summed up as “adaptive equifinality” in “just-in-time” performance. What was unacceptable in the HRO has become *sine qua non* for service and grid reliability in the HRN—but only in real time.

Several points must be noted. First, there have been experiments on the large scale—that is, on the scale of the California grid as a whole—which were grid-wide because the ISO could not do otherwise. They were undertaken involuntarily, that is, the ISO had little choice, for example, in introducing the proxy marketing software over the whole grid when it did. Over the course of the day it was introduced, complaints were made that price information was wrong, numbers were not showing up, and the information wasn't in real time.

Why experiment this way? Because the status quo had become untenable for the system operator, thanks to increased volatility introduced into the electricity system through the restructuring-induced crisis. BEEP stack bids were not coming in and operators had to do something. That “something” was to create proxy bids for the remainder of generation capacity

not bid in by the SCs. So too, the increased use of and reliance on the interties and imports to balance load and generation at the margins, "just-in-time," represents a grid-wide experiment that was in no way characteristic of the earlier integrated utilities. In these ways, real-time operations and experiments become synonyms.

The older HRO would never undertake such experiments, as it sought stable inputs and stable outcomes. In the HRN's case, there is no stable resting point for the network, because there are few routines and operating procedures that can stabilize inputs in order to ensure highly reliable outputs. Volatility always remains potentially high, when it is not actually high. One reason why the operations-as-experimentation can continue is that shedding load is a live option for maintaining real-time reliability. If "just-in-time" performance fails, operators can always shift into "just-for-now" or "just-this-way" performance modes. The increased legitimacy of real-time reliability has meant that shedding load no longer signals the end of reliability, as it did in the older HROs.

Improvisation-as-operations is occurring for at least four reasons other than that restructuring breeds churn and turbo-markets, rendering status quo untenable for any system operator. Technological innovation in reliability provision has been and continues to be a major driver of change. The electricity sector is notable for its search for better software, equipment, and approaches to real-time reliability management. A senior grid operations engineer at PG&E told us, "we now have a menu of analytic tools that we never had. You find out things that you never thought you would find out otherwise. We now operate closer to the margin, more economically, better utilization of equipment or installation of equipment."

Second, engineers are trained to innovate in the face of persisting problems (von Meier, 1999), and engineers dominate the ISO's wraparound. "Engineers always like to do new things," said one ISO transmission planner. Said an ISO engineering manager about his support work for the ISO control room, "It's pretty pioneering." Said yet another senior grid operations engineer, this time in PG&E, "I like pushing systems to the margins, but not one inch beyond. That's why we need to have that real-time capability, that knowledge of the system; having all the data coming in and the ability to crunch the data." We return to the importance of such professional norms in a moment.

Third, involuntary experiments are also occasions for learning, much as near misses are in other sectors. The real-time experiment, deliberate or inadvertent, becomes a design probe from which operators can learn more about the limits of service and grid reliability. "You don't learn as fast as you can, until you have to respond to something that requires fast response," argued a senior control room manager at PG&E's transmission operations center.

Fourth, the ISO has had many features of an organization that is in continual start-up mode with its double helix of incomplete design and hybridized markets. As with any start-up in its entrepreneurial phase, the ISO has been willing to try almost anything in order to achieve its objective of meeting the reliability requirement, all the time. "We created a new organization from scratch," said a senior engineer with ISO, one of the longest serving employees, "Part of it was trial and error learning. . . Part of it was our mantra of reliability through markets and you had to trust that it would work." And as with other start-ups, it has done so by pursuing

opportunities (i.e., taking risks) as and when they arise, especially in terms of adopting and adapting new technologies, software and communication systems. We call this, in terms of our framework, developing maximum and adaptive equifinality most characteristic of "just-in-case" and "just-in-time" performance conditions, respectively.

8. HRN Tolerance of What Would Otherwise be Considered Low Performance is Greatest in Real Time

No crisis is cost-effective. An earthquake happens or a humanitarian emergency occurs, and the first thought of those responding is not, Why couldn't this have been more cost-effective? Costs of course come into the calculus of decision somewhere along the line, but not in the press of the first response to do something, and to do it now before things get out of hand.

Clearly, there is little if anything cost-effective about "just-in-case," "just-in-time," "just-for-now" or "just-this-way" performance—unless, of course, you along with many others (the latest being the National Research Council, 2002) believe that avoiding the meltdown of our electricity infrastructure is the ultimate form of “cost-effectiveness”! That said, many energy economists, for example, would consider "just-in-case" performance to be too much like the older system of integrated utilities, where “high reserves” resembled excess capacity. If anything, the California electricity crisis has shown that we do not need operating reserves as high as the 13% of the older utilities or 15% at the start of the ISO. So why do we need them at 7% of "just-in-case" performance? “We don't really need 7% reserves,” said an ISO control room shift manager, “but we cannot do it every day like that!”

One could scarcely call "just-in-time" performance “efficient,” at least under current forms. Generators sell ancillary services that are not used by the ISO and end up as “overgeneration” on the grid, because they have to buy in bulk even when load is variable. One reason for this is the fact that scheduling is based on huge hourly blocks, whereas the load doesn't have such a step-wise change but a smooth shape. This discrepancy has even led one of our interviewees from the industry to develop new software to sell shapes rather than standard-sized blocks.

Other economic inefficiencies induced by "just-in-time" performance were also reported to us. We were told by a WSCC security coordinator that “everybody is in breach of their contracts,” including generators and the ISO, under the pressure of keeping things going at the last minute. Generation dispatchers regularly buy “too much” electricity, in decrements or increments, if CPS2 violations are any measure. “A lot of times the hour-ahead [market desk] over-procures, and the gen dispatcher works to get the estimates in line to what is needed,” explained a shift manager in ISO's control room. The ISO's double-paying generators—first in the day-ahead market for their electricity and in real time for not delivering it—was also fairly common under the pressure of the crisis. All this added expense—whether on the side of generators acting strategically or the ISO acting under mandate—is justified by the urgency of ensuring real-time service and grid reliability. “Last week we were paying them [BPA] to take our electricity, and now we're paying them to buy it. Why are we paying all the time?!” said an ISO scheduling coordinator.

"Just-for-now" performance is no better in terms of economic cost or efficiency. Here ISO officials go out of channel by calling generators or utilities directly, in some cases keeping generators online that may have been better offline in terms of their maintenance and longer-run service requirements. Yet "just-this-way" shedding of load has economic consequences that would be otherwise avoided were there any other alternative. They are still trying to tabulate the economic costs of the blackouts during the crisis, for example.

In short, the focal organization and along with it some HRN parties tolerate lower levels of performance, in strict economic terms, on the part of others as well as themselves, *if* those levels enhance or at least do not detract from their own overriding reliability objectives. In HROs, low performance was equivalent to increasing the potential for catastrophic cascading errors; in HRNs low performance, while wasteful, inefficient and at times outside the ISO tariff, is not catastrophic—indeed, it can help prevent catastrophe—precisely when it helps balance load and resources at the times it matters most, in real time. Of course, the current restructured system may well be more costly than the older system, but we cannot ignore the fact that neither system is cost-effective, at least to economists. We forget at our peril that controlling the stability of inputs and outputs was extremely expensive for the older HROs. All of this expense is the premium we pay and continue to pay for service and grid reliability in our critical infrastructures.

9. While Often Negative, Strategic Behavior Can Have Positive Effects on Real-Time Reliability

Strategic behavior on the part of the SCs and generators has been singled out by our ISO interviewees as the major reason why the electricity provision is not as reliable as it was perceived once to be. "Markets don't understand the problems of the grid," said one ISO scheduling coordinator. The withholding of generation to the last minute in order to increase profits has been well documented. Other reliability-threatening strategic behavior by SCs and generators was itemized by ISO interviewees, including "unscheduled flows," "uninstructed deviations," "not following dispatch orders," and holding up the day-ahead and hour-ahead markets so that bids can be made as late as possible.

There is, however, an important sense in which such strategic behavior, even though it threatens reliability in terms of WSCC standards, actually reinforced the importance of real-time reliability within the HRN, at least during the crisis.

It is often in real time that the huge profits were be made for a SC or generator, since that is where reliability matters the most during periods of shortages. It is moreover in real time where this strategic behavior has to be compensated, moderated or overcome by the system operator, since that is where it can best exercise its *de facto* or *de jure* role as buyer of last resort. That is, often in real time and only then on behalf of high reliability can both parties fully exercise their extraordinary powers—the SCs and generators their market power, the ISO its mandate as buyer of last resort. This coincidence of interests means that, while RRN standards of reliability are being undermined by strategic behavior of the SCs and generators, HRN real-time reliability would lose much of its compelling rationale without that strategic behavior—though presumably the HRN overall would be better off without the latter. Still, in a world were strategic behavior is

a fact of life, it does give a number of control room operators a job worth having and even enjoying. As one BEEPer put it, “A good day is when you’re both the cat and the mouse.”

Thus, strategic behavior may be zero-sum in some instances among parties within the HRN, and non-zero sum in others; each instance has to be looked at for determination. Nonetheless, the provision of reliable electricity is clearly less important “strategically” to the ISO’s longer term survival than it was to the older HROs, as the ISO’s future depends on many other factors than how well it does its job. If there is a lesson to be drawn from the PG&E bankruptcy, it is that its only branded product—reliable electricity—was not enough for it to survive in the market. As for the ISO, its organizational challenges come from outside the HRN, in the form of threats from the RRN and RTE, and not just from within the HRN. That said, PG&E is still carrying on its distribution role reliably, as far as we could tell, and we would expect system-operator functionality to persist, even if the ISO changes organizationally. Electricity is too important a critical infrastructure not to be treated reliably.

10. Market and Deregulation Incentives Exist for Ensuring Real-Time Reliability

Notwithstanding strategic behavior on the part of generators and traders, electricity markets and restructuring have also reinforced, if not in some cases improved, real-time reliability through a mix of incentives and disincentives that reinforce service and grid reliability.

With markets comes entrepreneurship and innovation. “Reliability Through Markets” has meant improved software and bidding systems for increasing the real-time bids in the BEEP stack. Strategic behavior of SCs and generators masked the fact that more generators were supplying more electricity to California than before restructuring, according to our interviewees. Since anything can make a difference at the margins when operating in real time, the importance of so-called “marginal” alternative energy sources, particularly wind energy, has increased significantly. “When it gets tight, the spot market prices go very volatile, and even small players can exercise market power,” said a major electricity advisor in state government.

There have been days when 700MW out of a total load of 35,000MW made the difference between blackouts and not, the figure being equal to the amount of wind energy being supplied in the state. In response, the ISO is trying to develop an innovative arrangement for scheduling wind energy under real-time conditions. As for blackouts almost all the economic and political incentives in the distribution system are to recover as fast as possible, since that is where most of the unplanned outages are and since the utilities now have the sole objective to do so (i.e., they no longer have what may have been competing goals to optimize in generation and transmission).

As we saw earlier in the report, market behavior comes in many forms and so too their real time. One example has been to contract out reliability-seeking behavior. This “contracting out” takes different forms. Some contracts directly stipulate requirements for reliability, such as in bilateral contracts between generators and distributors or RMR contracts. Some contracts are in the form of software, such as AGC, which enables the ISO to directly control generators it does not own but whose ancillary services it has contracted for. The AGC, along with RMR contracts, are contractual mechanisms between different HRN members that closest mimic the command and

control mechanisms between members of the older HROs. The proxy bidding system, which enables the BEEPer to “bid in” a generator’s unused capacity, allows the generator to have another chance to sell that capacity in real time—an option which in effect enables the ISO to contract out redundant generation that it may need in the last minutes before use. As it is, generators are required to provide some services “without compensation” as part of their contracted supply to the ISO, e.g., real-time voltage support.

Contracting out provides flexibility and predictability in volatile situations, like those of “just-in-time” performance. Instead of picking up the phone and telling the generator what to do as in the older HROs, the contracts and software stipulate how the similar results can be achieved otherwise and the options for doing so. In fact, the continual upgrades and drive for better software for the EMS, AGC and ADS systems is the HRN approximation of the HRO property of “constant search for improvements in reliability performance.”

11. Pulling New Participants into the HRN Becomes an Option for Increasing Real-Time Reliability

Clearly the most prominent proposal in this respect is that of instituting real-time residential metering. The consumer becomes part of the reliability equation. Metering’s rationale and relationship to reliability in real time have been explained a number of times. An economist, we imagine, might sum up the rationale as follows:

The point is that if you give consumers the ability to make decisions in real time, you would end up with the reliability *they* want. Reliability this way comes with a price. At present, reliability is buying energy at any price. If you want reliability “at any price,” then you obviously you will get it. But with the advent of energy markets, electricity is not about keeping the lights always on. It’s about consumers having the choice to consume or not. Having markets means now that reliability includes my decision not to consume. Consumers, like you and me, face a price and must choose and we can benefit or lose from that choice. Under the current system, consumers can’t manage that risk. Yet they manage precisely that kind of risk when they buy a car, an airplane ticket or stock and bonds; so why not power? Instead, regulators are protecting consumers and we are denying people these choices. That’s the argument for residential metering.

While there may well be these benefits to residential metering, we show in Chapter 10 that there are reliability costs as well, especially with respect to increasing system volatility facing the HRN in managing load and generation.

The attempt to bring customers into the HRN as a way of improving reliability covers more than just introducing real-time residential metering. Selected customers are currently brought directly into the HRN through the interruptible load programs that are triggered in stage emergency declarations as well as the induced energy conservation that takes place “automatically” when the different stages are declared. Such participation in effect bypasses the RRN and directly connects to the HRN, much as in e-commerce where the middle-person is cut out between the customer and supplier. As we have seen, when the margins are tight, customers reducing their load can be as important as a new generator coming online. Wind energy can be as important as a gas generator. The “local” RMR unit can have a grid-wide effect.

Generation has entered the HRN in other new ways as well, at least when the horizon goes no further than real time. There were, for example, aluminum plants that ceased production and shifted to selling their generation to the grid directly during the electricity crisis. In terms of the RRN, the effort to get FERC to establish and enforce hard, “contract-like” price caps has meant that FERC can become an important part of the HRN in a way it never was before the federalization of the California grid and electricity crisis.

12. Shared Professional Norms, Competence, Information and Accountability Substitute for the Culture of Reliability in Real Time

Much real-time behavior in the control room necessarily remains informal, tacit and unofficial in the face of enduring real-time pressures, and we have discussed how shared experience, knowledge and communication (such as cross-training and on-the-job training in operator positions) help ensure real-time grid and service reliability. It is in the control room where you see the HRO feature, a culture of reliability, persisting as best as it can in a networked setting. Restructuring however has ensured that the older culture can no longer be shared across the HRN as a whole. In our view, the most visible difference between the older HRO and the newer HRN is that the latter lacks a culture of reliability that the former had.

Strategic behavior and gaming by generators, it bears repeating, was the defining moment of the California electricity crisis. “Those guys in the generators, PhD in economics, they want to beat you at your own game. They say, just give me the rules,” argued a manager of real-time scheduling in the ISO, “When you do, they turn it on you.” “There is always someone who catches up to you and wants to play the system,” said an ISO grid resource coordinator. “Traditional market rules may not be applicable to [electricity]. They do not appear applicable,” concluded a state energy planner, “And that is disturbing. I think that you can devise any rules you want, there will always be somebody that will figure a way to get around the rules.”

Nonetheless, for all the gaming and strategic behavior contrary to the older reliability culture, the parties in the HRN have been able to partly compensate for the loss of the culture of reliability in at least four ways, all with real-time implications.

One response has been to accentuate the importance of professions and shared professional behavior across the HRN. Professional norms represent the thinking, values, methods and behavior dominant in any field or discipline. Sharing the same profession is manifestly important in ensuring real-time reliability within a HRN of divergent interests and open polarization, where some level of distrust is inevitable. One ISO senior operations engineer, as reported in an earlier chapter, told the story of how he and his counterpart in PG&E worked to avert a major power crisis in the San Francisco Bay area because they both were engineers with similar backgrounds. Most ISO control operators bring to their current positions the similar control room cultures of the former utilities. Indeed, many of the control room operators and senior staff in the wraparounds have actually worked with each other in the older utilities. There is even a sense in which restructuring can be seen as if the major integrated utilities had consolidated, reorganized and extended their control rooms by relocating existing staff in addition to hiring new ones.

Professional norms and learning curves carry with them blind spots and selective areas of inattention. We heard reports that engineers were writing procedures no operator could understand or that lawyers were proposing ambiguous language where engineers wanted no ambiguity. We also heard and encountered numerous differences between economists and engineers precisely over the issue of service and grid reliability in electricity supply. We heard an economist who said that in a properly run market there would be no need for operating reserves; when this was pointed out the next day to engineer in grid reliability he was at a loss for words. Doesn't he know that electrons obey the laws of physics, not economics? the engineer asked. In fact, the economist's point was that electricity now obeys the laws of economics.

We found differences between engineers and economists that extended well beyond their approach to operating reserves. Engineers we interviewed believe in simulations of large-scale changes prior to their implementation and those line operators we observed acted as if they subscribed to a version of the precautionary principle where large-scale changes are to be avoided until they have been shown, as best as possible under time and resource constraints, to do no or little probable harm. What if economists had been required to meet these standards prior to implementing their restructuring proposals? A theory that does not learn from its mistakes—i.e., the mistakes in California's electricity deregulation are because of bad politics and not economic theory, according to the economist's narrative—is not likely to place much weight on simulations or the precautionary principle, let alone high reliability.

Yet blind spots of narrow specialization can and do have a positive role in promoting real-time reliability. Without selective inattention, the professions would not have the problem solving capacities that they have. This is most clearly seen in the effect that a Stage 1 declaration has in activating and expanding the team of specialists in the control room and the fishbowl, when all kinds of wraparound staff come in to help, each with his or her own expertise and gestalt of experiential, institutional and representational knowledge bases. It may be more fruitful to think of selective inattention as "competent forgetting and competent remembering."

With professional norms and real-time competencies come "reliable databases" that are the lifeline of real-time reliability. The most remarkable feature of the control room is surely the one least remarked upon by our interviewees: There is not one operator in the control room that is not tightly linked to the outside through multiple communications and feedback systems. Everyone, all the time, uses the telephone; pagers going off all over the place; internal computers inside the control room "talk to" external computers outside; the AGC connects the generation dispatcher directly to the generators; the ADS connects the dispatcher directly to the bidder; dynamic scheduling controls out-of-state generators; governors on generators automatically bring frequency back into line; the frequency and ACE reflect real-time load and resources out there; all kinds of telemetry measurements come back to the control room in real time; web pages carry real-time prices and information; the day-ahead market operator has passwords to force balance SC schedules in real time; the security coordinator uses software to make the time error correction for the entire grid; control room monitors are nothing less than windows on grid reliability; and on and on. The control room for the focal organization of the California HRN is like a brain that exerts incredibly strong pressures for information to be reliable or, less metaphorically, for sharing information that is reliable because it contributes to real-time reliability.

As with a brain, misconnections sometime happen. First, there are software misfirings in real time. As one shift manager in the ISO's control room put it, "there is more pressure to put more RAS schemes in on top of RAS schemes to the point that. . . the potential exists for a large potential for cascading outages because we are operating so close to the edge with so many RAS schemes." Second, there is the gaming of economic strategies, counter-strategies and counter-counter strategies dependent on real-time information. Most prominently, real-time information has been used not only for the above-mentioned gaming of the system but also in other ways that threaten grid and service reliability. A senior grid operations engineer at PG&E explains,

We are not as close as when we owned the generators. Even those who are obligated to tell us are not all that forthright. . . That makes it more of a challenge. So there is a problem where you need to know the information ahead of time, but you don't want to know it because it could be used competitively, that is, other generators would like to know if there is a tube leak at Pittsburg. . . What was going on [in 2000] was so illogical, so convoluted, it wouldn't pass any test anywhere by those who know anything about power ops or economics too.

Restructuring, it must be stressed again, obliterated legal mandates to share certain kinds of information (i.e., generators do not need to share cost or generator outage information), while sharing real-time information between elements of the HRN has at times increased strategic behavior threatening reliability (e.g., the ISO pulled real-time information from its website that was being used by generators to game the market)—both of which reinforce real time as the *primus inter pares* of management horizons. Real-time information is so important that, when it is not there, is a major issue. "One of my problems," reported a security coordinator, "is that I don't have real-time access to scheduling information they [other security coordinators] have in Texas."

These problems aside, the overwhelming drive in the HRN is toward providing better real-time information, be it for improved forecasting, e-tagging software and EMS systems that allow real-time snapshots of the grid's current state. Sometimes the snapshots are from the computer, other times from the operator,

We want and are going to get an EMS system that allows you to take a real-time snapshot of the system that you can download and model with offline. Right now it is a manual job to take the model and tweak and tweak until you have something that you can use. (From our interview with a senior manager in the ISO operations engineering unit)

ISO has the ability to take a situation and evaluate it on a computer. We don't. I look at my screens and I got to make a decision. If we had a period of time where we could work with engineers and make an analysis of it and come back with some guidelines for giving assistance on real-time basis, that would be great. Basically, it is my decision. If we have the time, we would use the assistance; if we don't, we take a snapshot of system from what we know now, and look at what else we know. . . (From our interview with a shift supervisor in PG&E's transmission operations center)

Such representational, institutional and experiential snapshots are the real-time pulse-taking that increasingly substitutes for long-term forecasting. The "big picture" gives way to who has "the bubble" in real time or to what piece of software can give the operators "a snapshot" of the

system for "just-in-time" performance. "Everything is more and more complicated, pushing us into real time to expand our models," a senior manager in the ISO operations engineering unit told us. Said a senior grid operations engineer at PG&E, "I really believe you can't have enough real-time data. The system is running so incredibly close to the margins it's scary. After all, how long has the interconnected system been around? Not that long, so we hadn't been prepared for these kind of contingencies [such as a near voltage collapse in 1987]."

Part of the reliability of databases depends on error-admission and information sharing, both of which are important to real time. Each weekday at 9 a.m. a meeting is held in the ISO's fishbowl to share information about the day ahead and other matters. Information sharing was routinely observed by us during our control room observations. We also observed the voluntary admission of error and accepting fault for mistakes, which is another HRO property. It may be these misjudgments are unambiguous and obvious to everybody because they most often concern the balancing of load and resources, which accounts for them being so readily admitted.

Nor are real time and long-term necessarily incompatible. In one sense, the always-on feature of electricity is the long-term version of real time. Longer-term planning and investment in the grid are increasingly dependent on the real-time data, i.e., planners need to be constantly updated on real-time demand, supply and system states and changes. What parties in the HRN have discovered is that critical information must be distributed and shared throughout the network in real time for purposes other than real time. Real-time information turns out to be crucial for longer-term issues affecting service and grid reliability. Logging of every action by the BEEPer and gen dispatcher are, for example, reviewed in meetings and by wraparound staff in order to detect and anticipate problems affecting reliability of the grid.

Finally, while the HRO culture of reliability has been attenuated in the HRN and appears non-existent in some important HRN parties, informal accountability exists and persists for behavior in real time, when balancing load and generation has its greatest urgency. Because of the always-on nature of electricity as a service, participants in the HRN frequently know when network performance conditions are changing. The pace picks up, calls get more urgent, and by some point everybody knows, including customers, that the service is in jeopardy. The generator knows when a phone call from the ISO has to be taken seriously, that is, when the ISO is in the midst of "just-for-now" performance. Accordingly, knowledge of whom the control room operator can trust or has to distrust emerges out of the give and take of maintaining the balance between load and generation.

This knowledge as well as the trust itself represent a structure of informal accountability within the HRN that can buffer against volatility and increase options variety in other ways. The accountability enables one to be judged over whether or not what he or she is doing helps maintain real-time reliability. Unlike the culture of reliability, the web of informal accountability is based on self-interested behavior in ensuring the always-on service—because if it is not always on, every one's self interest suffers. Informal accountability rather than a culture of reliability substitutes the HRO preoccupation with respect to safety with the trusts and distrusts of HRN operators. Just as restructuring has ensured that the future is never so reliable as it now, so too is the reliability network created by restructuring never as trustworthy as it is in real time.

The pressure will undoubtedly be to formalize these informal accountability structures. Presumably, the more accountable you are for something, the more responsible are you for it; and the more accountable and responsible, the more the pressure to formalize that accountability and underlying responsibility. Our report, however, has given ample evidence to question this assumption and fear for its implications.

Designers must be chary about how much of "just-in-time" performance and associated workaround behavior can be formalized, given volatility that cannot be controlled or predicted and given the adaptive, last-minute nature of the equifinality involved. The balance between adapting to the unpredictable yet having formal rules in case the worse becomes unavoidable is reflected in remarks of a PG&E shift supervisor,

Out of the six [blackouts], I've done four. . . It's been a changing, living document. What's been happening has never been done. So we're trying to write rules for something that's just happened, but each rolling blackout is new. It was a nightmare.

Front-line discretion and back-office rules do not mix well when the mandate is to avoid the train wreck that has never happened before. From the perspective of our report's framework, any formalization of accountability must reduce system volatility, increase network options variety, and/or enhance the cross-performance adaptability of the focal organization to respond to changing conditions in both. A history of the ISO would be one measured out by designs that failed this very test. In explaining why an early version of ISO software failed, a WSCC security coordinator told us, "This was because there were no real-time people involved in the design. They thought that hour-ahead bids would be final and the system had no capability to change in real time."

13. Reliance on Real-Time Information and Software Leads to Reliability in Real Time

In many systems, a gradient stretches between use and management, such that at one end of the gradient you find that the way people use the system is the way they actually manage it while at the other end users and managers are completely separate groups. You see the phenomenon of use-as-management everywhere. In Africa, women collect water from a fenced dam by climbing over a stile into the dam area, i.e., the way they use the dam is the way they manage to keep polluting cattle and small stock out. In Asia, people walk back to their village on paths following the irrigation canals, i.e., the way they use these paths is the way they manage to continually oversee if there is anything wrong with the canals. In the greater Bay Area, roads circle around the top of Delta islands, which afford their drivers an easy view of what is happening to the agriculture on one side and the Delta watercourse on the other. Remember Napster? There too, the reliability of Napster was a function of the number of people who actually relied and used it. There is also an element of this use-as-management in real-time reliability.

We have mentioned the continual reliance of control room operators on their software systems, such the ADS, AGC, and EMS. We have also underscored that much of this software and software development is geared to generating real-time information and that operators increasingly rely on that information to balance load and generation hour by hour until the last

minute. Thus, looking across the HRN, the more that the ISO, utilities and generators rely on real-time information from their software systems, the more the service will be reliable in the way it is, i.e., through "just-in-time" performance. The more operators have to rely on new or emerging reliability criteria, the more the service will be reliable in ways that depend on the ISO having the adaptability to move between different performance conditions. Indeed, *the way ISO operators actually use and rely on their software*—e.g., manually dispatching generators in the AGC, phoning generators who do not respond to the ADS, relying on basic system measurements of frequency when the EMS system is having problems yet still relying on the AGC, ADS and EMS all the other times—is part and parcel of the way the service is being managed for real-time reliability.

As philosophers counsel, the meaning of something lies in its actual use, and never more so than for those pragmatists par excellence, the control room operators. Their reliance on real-time information from most of the software begets "just-in-time" performance and vice versa. Such positive reinforcement of real-time reliability is one of the strongest pull factors we observed.

14. Flexible Authority Patterns and Teams Become the Locus of Reliable Behavior

The report has spent considerable time on discussing the wraparound and its importance for "just-in-time" and "just-for-now" performance. Therein lies its importance for real-time reliability.

As in HROs, the widening teams (in our case, within the control rooms and across the HRN) respond to and cope with the faster pace and urgency through resorting to flexible authority patterns based on experience, skill and discretion, rather than role, procedure, and protocol. Teams combining different positions are a matrix that goes lateral or vertical, depending on what is required in real time: One minute, the shift manager is treated as colleague whose advice matters to the operators, the other minute s/he is called up to make a decision or buffer them from outside interference. The teams are able to take on a matrix structure—the HRO property of flexible authority patterns—precisely because hierarchy gives way to experience and skill when real-time reliability is at stake.

But we should never lose sight that the matrix works precisely because going lateral carries the possibility of going vertical in order to ensure grid and service reliability. “Sometimes I talk to the generation owners,” an ISO shift manager observed, “but only if they don’t do what we need them to do [in order to ensure grid reliability]. Bully them if needed. . . . Sometimes I pull a little rank, a little authority.”

15. HRN Relations with the RRN and RTE Seek to Foster and Protect Real-Time Reliability

How HRN parties manage their relations with the RRN and RTE affects their real-time management of grid and service reliability. The impact of the California electricity crisis on the system operator, while making things difficult, could have made the state of affairs much more

complex for the HRN than it actually did. In terms of our framework, the unicentric focus of the ISO is on the achieving the reliability requirement through four sets of performance conditions that vary in terms of system volatility and options variety. The consequence, in doing so, has been to shift the complexity of intermediate and longer-term issues to the RRN and RTE, leaving the ISO with its more urgent problem of ensuring real-time reliability.

The harder issues of deciding whether the state should buy the SCE grid, at what level to fund the new power authority, change the ISO governance structure, or extend CERS have been left to the RRN and RTE, issues that the ISO control room and wraparound staff have avoided—and with good reason. At the mid-year height of the crisis in 2001, a very senior ISO official told his staff, “My view is that we [in the ISO] are jumping under the table and the earthquake is happening and what we have to do is to hope this isn’t a nuclear attack, and that the rubble will settle and when it does, we get up and will be the only ones around who know what to do. We’re a third world country now and the \$13 billion is now in Houston that would have cushioned us from the drop in the [state] economy that is going to happen.”

16. Unbundled Generation Can Increase Predictability, Transparency and Self-Organization in Real Time

The profit motive of generators has been treated as if it were acid thrown into the face of service and grid reliability. Yet there is a sense in which unbundled generation increases predictability and self-organization. Under restructuring, the generators have a more focused mandate than the utility-owned generation before them. The concern of deregulated generators in terms of reliability is not that of the grid, but of “commercial availability” and meeting their energy commitments. Not meeting those commitments has direct financial implications to which they are very sensitive and for which they are accountable. Consequently, generators self-organize to deal with contingencies so as to not be exposed to the financial and other consequences of not meeting their commitments.

To that end, they have increasingly adopted the major innovations and instruments of the financial sector, including hedging, portfolio management, options trading, and risk management. Much has been written about the problems of these instruments. Yet one of their advantages deserves underscoring: The ISO and distribution utilities no longer need to know how the block of contracted energy was assembled, at least in the same way the utilities had to in the past, as long as the electricity is there when it matters. The transaction costs of generating and assembling contracted energy are no longer spread across the HRO but located in the decoupled, self-organized activities of one set of HRN participants. In the words of a senior generation executive of a state energy supplier, “If we can’t meet our scheduled commitments, we have to purchase what we’re short at the real-time price.”

Similarly, many of the recent events over abuses of Enron *et al* during the California electricity crisis obscure one of the major advantages of the restructuring: Behavior affecting real-time grid and service reliability is much more transparent than it was ever before. It was abundantly clear to the operators we interviewed that strategic behavior and gaming on the part of generators was going on during the electricity crisis at the time we were talking to them. What has been treated by FERC as the “revelations” of the Enron memos in May 2002 were repeatedly reported to us in

our mid-2001 interviewees. The October 2002 admission a top Enron trader to undertaking practices that manipulated market prices during the California electricity crisis added nothing new to what was already generally known in the control room a year and a half earlier.

Such observations were possible primarily, we believe, because of the increased transparency that the marketplace brings to transactions that had been undertaken within “the black box” of the integrated utilities and their regulators before. “The balancing of cost and reliability use to take place in the black box, but with deregulation you are trying to take what was complex and Byzantine in the black box and drive it by economics,” said a former PG&E executive, “Now, [this has] blown the black box apart.”

The market has made important (but not all!) transactions more visible than ever before. Before we didn’t know what was happening inside the integrated utilities, save what the CPUC was able to find out; now withholding energy to the last minute, and breaking contracts to do so as to generate returns that more than compensate for penalty charges, is an exercise of market power clear to insiders in the HRN and RRN, if not beyond. One of the more piquant ironies of the California electricity crisis was hearing pro-restructuring pundits insist that nothing like market power was being demonstrated by the crisis—it was really the climate, population increase, price of natural gas, environmental overregulation, flawed deregulation—when in fact restructuring’s greatest success was to create a marketplace in which it was patently transparent to its buyers and sellers that such market power did exist and was being exercised. “When you see it, you believe it, if you’re an operator,” concluded one of our interviewees.

Conclusion

At the start of Chapter 2, we said the most interesting feature about the California electricity crisis is that no one predicted it. Now we know why even insiders—especially insiders—failed to see it coming.

For surely part of the reason is the rise and preoccupation of operators with real time. If you can’t predict hour to hour, how can you predict what would have happened months ahead or years beyond? Clearly a significant portion of the real-time reality is the difficulty to forecast, which is, in turn, another version of the confusion and causal incomprehension found in “just-in-time” performance. “How can you predict without a longer-term perspective?” is precisely the question we found missing and absent in the HRN control rooms and wraparounds we observed.

That is real time's downside. It should go without saying that each push and pull factor to real time carries with it risks, costs and disadvantages potentially invidious to the long-run provision of reliable electricity to Californians.

Still, if we are right in our findings and argument, the need to balance load and generation under high volatile system conditions is likely to persist indefinitely, whatever transmission and investment efforts now undertaken. We have outlined sixteen good reasons (there are probably more) why this expectation of real-time reliability is a sound one—but one that is never guaranteed and only realized hour by hour with little to spare. Such a conclusion may leave some readers in consternation: “But there is no guarantee adjacent control areas will cough up the

megawatts needed at the last minute. Discretion is all good and fine, but we still need rules to ensure no abuse is happening as a result. How dare they experiment on us without our knowledge! Are you saying that we should be tolerating all this low performance—buying energy that we don't use, paying generators twice, deferred maintenance and transmission investment, and more? Who's talking about what we are doing to the environment through all these wasteful purchases of electricity?" And so on.

May we be so bold as to submit that these and like points are irrelevant to the issue that this report is bringing to light and underscoring. Such questions have zero—repeat, zero—policy relevance in the absence of the reader being able to provide ways to change those conditions that push and pull operators to the left side of our framework, that is, to "just-in-time" and "just-for-now" performance—all in the name of the service and grid reliability that the reader expects, often without question and often treated as a given.

For this long chapter has been devoted not to just one or two factors that reinforce real-time reliability, but many, many factors. Granted, no one factor is sufficient to compel real-time performance, but there are numerous other reasons to be performing in real time. Today's real-time is an alternative version of what will happen tomorrow instead. The only place the future is reliable is now. The policy issue facing Californians today—the hitherto untold story about the electricity crisis—is that the pressures pushing and pulling elements of the HRN to realizing reliability only in the very short run are, quite simply, quite sadly, over-determined. We suspect that quite many angry Californians understand this intuitively.

What does this mean, policy-wise? Baldly stated: Increase generation, and we will still see real-time reliability, not as frequently as in the crisis, but regularly enough to realize that more generation alone is not the answer. Improve Path 15 and we will still have real-time reliability because there are always other “weak-points” in need of immediate remedy. If there is not one reason to wait until the current hour or last minute to keep the lights on, there are many more reasons to do so gathering off-stage in the wings. The good news, however, is that these sixteen factors incorporate many, if not most of the features one would want to see in complex organizations and networks that must adapt to a rapidly changing, unpredictable task environments in order to achieve reliability mandates that society and the public at large do not want compromised. To see what the officials could be doing better, we turn now to the recommendations in Chapter 10.

10

RECOMMENDATIONS

If the people of California are to draw from the California restructuring and crisis any useful lessons—and so many wrong ones are now being drawn for them—we must know what actually worked to keep electricity on when so much was working to turn it off. The preceding chapters have been a modest step in that explanation.

There are many challenges facing policy makers, but the ones we focus on below are related specifically to the report’s focus on real-time performance constraints on service and grid reliability. The subsequent sections summarize our recommendations which follow from the report’s four major findings:

- Reliability issues were not a significant enough part of the debate and policy design process surrounding restructuring in California. The grid was treated essentially as a “black box” in the analysis of players and their market relationships. The full transaction costs of market exchanges were not and still are not factored into pricing and rates because the reliability burden imposed by these exchanges is not internalized by the market players. Yet service and grid reliability are thoroughly taken for granted in the calculations and market strategies of generators and end users. It is important to note for the future strategy and expectations of market players and regulators that the one-time adage of the restructuring “reliability through markets” should actually have been “markets through reliability.”
- The major professions that informed policy debate and design over restructuring—economics, engineering, politics and policy analysis—failed to bring reliability issues into proper perspective. Economic analysis treated reliability management as a free good for market participants. Engineers failed to address the organizational and institutional dimensions of providing reliability. Policy analysts failed to appraise policy makers and the public of just how radical a departure from the representational, institutional and experiential knowledge bases restructuring was and the potential for displacement of the risks onto the public of this open-ended policy experiment. Finally, as near as we can determine, managerial perspectives were conspicuously absent from the design of the ISO and the PX. There was no appraisal of the organizational and administrative challenges and role conflicts that would be likely throughout the restructured system in working in such an untested and unstable environment.
- Because reliability has been and continues to be the neglected perspective in the full gambit of electricity system design, special efforts must be made to incorporate reliability issues into future changes, both in policy and in practice.
- While much attention of the ISO and its regulators focus deservedly on improving the electricity system so as to avoid blackouts or, barring that, to ensure recovery from blackouts as quickly and safe as possible, such attention focuses only on the right side of our

framework, i.e., "just-in-case" and "just-this-way" performance modes, where system volatility is low. Our research indicates that, notwithstanding these improvements (and sometimes precisely because of them), fits of high system volatility and uncertain options variety must be expected in the future as a way of operating life and addressed with better foresight. This means much more attention should be given to improving performance on the left side of the framework in terms of "just-in-time" and "just-for-now" performance.

The Challenge of Reliability

The recommendations based on our findings are tempered by the cautionary perspective regarding policy design and redesign that follows from the analysis of the preceding chapters. It is all too easy to offer prescriptions, harder to implement them, and often harder still to live with their consequences.

We would be foolish to set out major recommendations, as “The ISO should reengineer its processes for this...,” “The CEC should revamp its processes for that. . .” or “The Governor’s Office should. . .” A close reader of the report knows by this point that major design initiatives got Californians into the electricity crisis and major overhauls undertaken in the same way are likely to make matters worse. Many policy proposals are currently being made to change the electricity system. The focus of our recommendations is different. We emphasize instead the professionals responsible for reliability management, not the design problems facing further restructuring, whether regulated or deregulated. We do so because the real story of the California electricity crisis is not about structures, but the skills that matter under changing performance conditions, especially real time.

Our starting point is where Chapters 3-9 leave us: We identified a set of reliability professionals in the high reliability network, but we also found no formal professions that embrace the knowledge base and skills required for this reliability under networked settings.

The reliability professionals we observed and interviewed know how to keep the lights on under adverse, extenuating performance conditions. The professions of economics, engineering, policy analysis, law and others, which are so often applied to the design and operation of critical infrastructures, do not as yet understand nor appreciate the full real-time reliability requirements of these infrastructures. While many specialists in these professions deal with reliability issues, they are not reliability professionals. These specialists most certainly do not focus on the critical balance between options, volatility and adaptability that we have highlighted in this report.

The formal professions participating in the California restructuring exercise sought to design and implement something that was beyond the knowledge base of their disciplines and fields. We saw social engineering by professions whose members do not even think of themselves as social engineers. They have been engaged in single discipline-driven design of a multi-disciplinary system. What was new and at the intersection of many disciplines was treated as if it were a matter of straightforward market, engineering or political design. What was ignored, however, was the heart of the matter: Economics, engineering and politics are not theories of high reliability. You may get a service and grid reliability from the three professions, but never all the

time nor as the paramount priority. For that, you need reliability professionals and their professionalism.

There are many people trained in the individual disciplines who are working as operators, analysts, economists, engineers, and programmers in the HRN and who are in actuality first and foremost reliability professionals. They are professionals in an emerging profession that is as-yet unrecognized. They have knowledge bases and professional roles in the management of the electricity system that all of us need to better recognize and comprehend. Our recommendations are focused on the challenges these reliability professionals face in network management, the challenges to developing their emerging profession and communicating their skills, and the challenges to us, as Californians, in taking reliability and its professionalism around it seriously.

The Neglected Professionals

The challenges regarding grid and service reliability in the critical infrastructure for electricity are best understood within the space of action available for high reliability management.

That space can be characterized along two dimensions: (1) the type of knowledge brought to bear on efforts to make an operation or system reliable, and (2) the focus of attention or scope of these reliability efforts. The knowledge base from which grid and service reliability are pursued can range from representational knowledge, in which key activities are understood through formal principles and deductive models based upon these principles, to experience, based on informal or tacit understanding, generally derived from trial and error. By embracing perspectives that span this range, the knowledge of a reliability professional overlaps but is not congruent with the knowledge base of his or her individual, established discipline.

At the same time, the scope of attention of the reliability professional can range from a purview which embraces reliability as an entire system output, encompassing many variables and elements of a process associated with producing a stream of reliable results, to a case-by-case focus in which each particular event or process has distinct properties or characteristics. These two continua of knowledge and scope define the space within which reliability can be pursued by reliability professionals (Figure 10-1)

At the extreme of both system-wide scope and formal principles is the design approach to reliability. Here formal deductive principles of an individual profession—e.g., economics, law, engineering—are applied to understanding a variety of critical systemic processes. In this corner of the world, it is considered inappropriate to operate beyond the logic of design analysis, and analysis is meant to cover an entire reliability system, including every case that matters to which that system can be subject. Design in this sense is more than analysis; it is major control instrument for system behavior. This approach, for example, dominates the nuclear power industry where operating “outside of analysis” is a major regulatory violation. It is also evident in the Stage 3 emergency load-shedding under which “just-this-way” performance is mandated.

At the other extreme—single or case-by-case focus and experience-based understanding—is the activity of constant reactive behavior in the face of case-by-case challenges. Here reliability

resides in the reaction time of control room operators rather than the anticipation of system designers.

		Scope of Reliability Focus	
		System-Wide (All cases)	Specific Event (Single case)
Knowledge Base	Representational (Formal, deductive)	DESIGN	Contingency Scenarios
	Experiential (Tacit; trial and error)	Strategic Adaptations	REACTIVE OPERATIONS

Figure 10-1
Reliability Space and Key Professional Activities

Both positions are, however, extremes. Each alone is insufficient as an approach to providing grid and service reliability, though each is necessary. Designers cannot foresee everything, and the more “complete” a logic of design principles attempts to be, the more likely it is that the full set will contain two or more principles which contradict each other. On the other hand, case-by-case reactions by their very nature are likely to give their holders too specific and individualized a picture, losing sight of the proverbial forest for the trees. Experience can become trained incapacities that lead to actions undermining reliability because operators may not be aware of the wider ramifications of what they are doing.

What to do? First, “moving horizontally” across the reliability space directly from one corner across to the opposite corner (i.e., upper left to upper right, lower right to lower left) is likely to be unsuccessful. A great deal of reliability research supports the findings of this report to the effect that attempts to impose large-scale formal designs directly onto an individual case—to attempt to anticipate and fully deduce and determine the behavior of each instance—are freighted with risk (Turner 1978; Perrow 1986). Yet this is what system designers attempted at

times to do to secure reliability in the California system—to, as one engineer described it, “design systems that are not only fool proof but damned fool proof.”

Instead of horizontal, corner-to-corner movements, Figure 10-1 shows that reliability is enhanced when shifts in scope are accompanied by shifts in the knowledge base. Given the limitations of the extremes in this reliability space, it is important to operate in positions closer to a shared center both by moving from reactions to building patterned recognitions and strategic adaptations and by moving from designs to contingency planning and scenario building. It is difficult to find this middle ground, to combine macro and micro perspectives on a complex system and to bring together logic and experience and (perhaps even more difficult) designers and practitioners. It is, however, in this middle ground where experience, as we have seen, is trusted over doctrine, discretion over design, and a high degree of improvisation and inventiveness is often applied.

Indeed, the "just-in-time" performance we observed within the control room and wraparound of the system operator, where the immediate marginal balance between load and generation is center-stage, falls to the professionals who can master operations at this level or within this portion of the reliability space. It is important to recognize and build upon what has been invisible to outsiders—namely, this group possessing unusual skills, knowledge and commitment to the reliable operation of the electricity system. It is with this group and their professionalism where we believe the greatest gains in grid and service reliability are to be found. The future of reliability in this system, if you will, lies in making the invisible visible and the visible a priority. Reliability professionals must evolve in terms of promoting new recruits, professional careers and analytic capacities. We turn to the challenges now.

Building Analytic Capacity

A common move in reports is to recommend “increased capacity building” and then to proceed to identify what specific units or staff should undertake that responsibility. We in fact have done that ourselves in other reports. Here, though, our focus is on the more useful point, Just what “analytic capacity” needs to be developed and established? What does, for lack of a better term, “reliability analysis,” look like? (We keep the quotes around the term in order to distinguish our version of "reliability analysis" from the more common variants in engineering and industry.)

Our report has at several points reiterated the need to analyze any proposed improvement in the California electricity system in terms of whether its implementation would increase system volatility, decrease the variety of options to balance load and generation in response to that volatility, or interfere with the cross-performance adaptability (or maneuverability if in “just-for-now” performance) of those responsible to maintain and ensure grid and service reliability under changing conditions of system volatility and options variety. A proposal must pass the reliability-matters test. If the analysis shows that the balance of these three issues is negative, the proposal should not be implemented. Reliability professionals, again in our view, are better able to assess and balance options variety, system volatility and the adaptability between modes of performance conditions. If the California electricity crisis proves anything, making that balance is not really possible at the extremes of the reliability space in Figure 10-1. It is exactly this balance which has been ignored in the design exercises of the individual disciplines and professions that contributed to electricity restructuring. The balance is also what has been lost in

the preoccupation with real-time operations under conditions of high volatility induced by the restructuring. How then do we analyze our way to the middle ground where this balance can best be made?

The Way to Go

The reliability professionals who were very good at taking reliability seriously under adverse conditions are those who combine the extremes by first searching out patterns and scenarios.

Return to Figure 10-1. The best way toward the middle ground requires those at the two extremes of design and real-time operations to move obliquely toward one another, as if tacking in the wind in order to get to one's destination. More formally, "reliability analysis" requires looking for pattern recognition across multiple experiences as well as for scenarios generated by thinking through design logics when applied to multiple contingencies and cases.

We know from research and observation that reliability is enhanced when designers apply their designs less globally, while at the same time relaxing their commitment to a set of principles that fully determine system operations. This happens when designers embrace a wider set of contingencies in their analyses and entertain alternate scenarios for system behavior and performance. If the design logic says $a \rightarrow b \rightarrow c$, can applied designers imagine contingencies or scenarios arising which would lead to the exact opposite? This process usually leads to the formulation of emergency or other response protocols as well as standard operating procedures which handle an enhanced range of possible system conditions.

From the other direction, reactive operations can shift away from firefighting toward recognizing and evaluating patterns across a run of real-time cases. In this way, managerial adaptation—recognizing common patterns and evolving norms, strategies and even routines to cover similar categories of events or cases—arises. These emerging norms, strategies and routines are likely to be less formal than the protocols developed through contingency analysis or scenario-building.

The preceding chapters provide robust evidence of both oblique approaches to the middle. New nomograms are being developed on the basis of contingency analyses by operations engineers in the ISO. New marketing and scheduling routines after the collapse of the PX continue to emerge and adapt through the trial and error learning of the ISO market engineers. The more important point here, however, is that reliability professionals are better able than those at the extremes of the reliability space to pull the pieces together toward the middle—formal design, real-time operations, pattern recognition, and contingency scenarios—when it comes to ensuring and safeguarding reliability. In particular, reliability professionals understand better than many that the burden of proof should be on those who propose design changes to make compelling arguments that a change will not negatively affect the net balance between options, volatility and adaptability.

Three Examples of "Reliability Analysis"

Terms, like “balance” and “pulling together” are all fine and good, but how would reliability professionals analyze an issue, if they were operating in middle ground of the reliability space? If professionals had the analytic capacity, how would it actually work?

We choose three examples, the first illustrative, the second and third because of their topicality. Our format in each is to sketch how a "reliability analysis" can be undertaken. Because of space limitations, nothing that follows pretends to be a full-fledged “reliability analysis” of each example.

Stage 3 Declarations. A joint report of the California Electricity Oversight Board and the California Public Utilities Commission (Kahn & Lynch: 2000, vii) recommends “ten potential actions to prepare for an electricity emergency,” the second of which is: “Authorize the California Public Utilities Commission working with the utilities to determine when to shut off electricity in a Stage 3 emergency.” Again, we chose this example not because of any special importance it may have, but because of the points it illustrates.

A "reliability analysis" of this proposal would proceed by going beyond initial design and real-time operational considerations. For example, the proposal, if implemented, would likely undermine network reliability by diminishing options variety and system operator adaptability during real-time operations. The option to declare a Stage 3 is currently made inside the ISO control room by people on the spot; the proposal would take that decision outside the control room to be made by people who do not have the “bubble,” even if they were to have access to immediate, real-time information (which itself is doubtful). The result would be an increase in errors when trying to balance load and generation at precisely the moment it matters most in terms of service reliability’s threat to grid reliability. The ability of the ISO to shift and adapt across performance conditions would be reduced, since it no longer has the option on its own to shift in a timely way from "just-for-now" to "just-this-way" performance.

If such a proposal were implemented, a plausible scenario would be that the system operator would no longer push the limits under "just-for-now" performance, knowing that it no longer had the option to move quickly into "just-this-way" performance at its own discretion. That option would now reside in the CPUC. This would mean that well before those limits are reached, the ISO would likely turn the decision over to the CPUC (or any other outside organization charged with this decision), saying that it can no longer guarantee grid and service reliability, thereby putting pressure on the CPUC to move the system sooner into blackouts or accept possibly losing control over the system and risking collapse. The CPUC would be expected to counter-game, but ultimately it does not have the control room experiential knowledge base of pattern recognition and adaptation from which to argue otherwise with the ISO. Accordingly, we would expect the network to be less reliable, were such a proposal to be adopted.

Real-time Residential Metering. Our second example of "reliability analysis" is the more topical proposal for real-time residential metering. Start by thinking in terms of design and real-time operations. Consider the fierce design assumptions of many economists and engineers with regard to real-time residential metering. They ask us to: Assume that real-time metering for all

electricity customers can be implemented. Assume that the meter technology has been thoroughly tested before being introduced, i.e., there are no computer or software glitches like those plaguing the ISO since its inception. Assume that there will be no common-mode failure in the technology, e.g., meters going out all over southern California, thus transforming the 2001 California electricity crisis into a dress rehearsal for what is likely to happen should the meters fail on a large scale for an indefinite period. Assume as well that all customers have real-time pricing information and can make "just-in-time" adjustments in light of these price signals. Assume, in other words, that none of the preceding is a problem. Assume all that has happened without noticeably increasing system volatility in any noteworthy way.

System volatility might well increase with residential metering, notwithstanding the above assumptions, precisely because real-time residential metering introduces substantial uncertainty into ISO pattern recognition and contingency scenarios. Most prominently, residential metering would bring elements of the RTE directly into the system operator's control room. Under the current system, the ISO deals with highly aggregated consumers through the intermediary of the distribution utilities. They coordinate the activities of their customers, and through their learning and experience with customer behavior, have made load forecasting more predictable. The utilities also centralize the load shedding process if required under the Stage 3 emergency declaration. Without these coordinating units, the ISO would have to fashion load forecasts based on the anticipated behavior of large number of discrete, metered consumers and would have to interact with them directly in order to effect load shedding under Stage 3 conditions. Such contingencies—namely, disrupted pattern recognition upon which to base reliable load forecasts—are likely to introduce unpredictability and uncontrollability—namely, additional system volatility—into the ISO's ability to meet its reliability requirement.

A "reliability analysis" might well conclude that real-time residential metering could reduce the variety of network options through any number of different scenarios. One scenario is that when the ISO calls up more generation, it now faces suppliers who need more finely-grained information in order to make decisions about whether it is worth offering that supply. In a system with large distribution utilities more coarsely-grained information is readily available, because it is aggregated and more certain in light of the long experience of the distributors with aggregate demand (i.e., load forecasts are based on proven pattern recognition and contingency scenarios). In a new residential metered arrangement, more and different information will be needed in order to determine whether providing additional supply is financially worthwhile to the supplier. In the absence of that information, some suppliers may adopt a wait-and-see strategy before investing in or offering new supplies to the ISO. If so, this supplier strategy could limit the options available to the ISO. Without new software for better forecasting based on discrete metering and pattern recognition, the use of available network options might also be more error-prone. For example, the use of AGC is highly tied to the predictability of load forecasts.

Other questions about residential metering are raised once we move beyond design, and real-time operations consideration to contingency scenarios and pattern recognition. As economists remind us, with metering, consumers would now set their own "reliability" standards for residential electrical use. Would this impose different criteria for reliable provision of services on the ISO? Would these consumers, for example, accept the same pattern of billing errors in their electricity bills as they currently experience in their telephone or credit card bills, knowing that the former

errors could wipe out a whole year's worth of energy conservation savings? Would the ISO now have a call desk to take the calls of disgruntled residential consumers? How many different languages would the bills have to be sent out in or the call desk be able to field? For that matter, just what level of education and training would be needed for consumers in residential metering for this minority majority state? In short, it is an understatement to say that the design assumptions and real-time operational requirements of residential metering pose formidable challenges to contingency management and organizational adaptation within a networked reliability setting.

Homeland Security and the Electricity Grid. A momentous debate is taking place among policy makers and technical experts over how best to protect our critical infrastructures against attack. Particular attention has been given to how better to protect the electricity grid (Farrell et al 2002; NRC 2002, see also Homer-Dixon 2002; Mann 2002). At issue is the notion that critical infrastructures have discrete points of vulnerability or "choke points." If they are attacked or fail, the performance of the entire infrastructure is threatened. Key transmission lines upon which an electrical grid depends, or single security screening portals for airline passengers, or a financial audit on which countless investment decisions are predicated become points of vulnerability which can lead to cascading failures well beyond the system in question.

From this viewpoint, the remedy is to redesign complex systems so as to decentralize or decouple the parts and render them less interdependent. Decentralize power grids and generators, we are being told, making smaller, more self-contained transmission and distribution systems. Indeed the existence of choke points, it has been urged, signals the vulnerability of these systems, both inviting and guiding terrorist attack.

A "reliability analysis" and the experience of the reliability professionals in the restructured California grid during the 2000 – 2001 energy crisis question this remedy and suggest that many proposed redesigns could well undermine the very features that ensure operational reliability. Their analysis would begin with pattern recognition and contingency scenarios. For the reliability professionals who manage the California grid, choke points can be a system resource as well as a source of vulnerability. Despite their attractiveness to terrorists, choke points are also the places reliability professionals give most of their attention.

Tightly-coupled choke points allow reliability professionals in the control rooms to see the same portion of the system as the terrorist and allow them to assess operations as a whole against a backdrop of alternatives, options, improvisations and counter-strategies. Decentralized systems present many more independent targets for sabotage. Decentralized managers would also probably not have a clear picture of what is happening overall, nor will they have as wide a range of alternatives and recovery options. While malfunctions may not bring down major portions of a decentralized grid, attacks anywhere can give the impression that the grid is now everywhere open to new risks.

The scenario of complex technologies cascading out of control under precisely-targeted destruction by terrorists is certainly not one that any of us should lightly dismiss. Yet these technical systems have always had the capacity to "sabotage" themselves, just as outsiders have had the ability to assault them strategically, as some out-of-state energy suppliers and traders did

during the California electricity crisis. Indeed, it is difficult to imagine terrorists threatening the state's electric grid more than we have in the last few years.

Do these "reliability analyses" mean there should be no residential metering or no decentralization of the grid whatsoever? Of course not. It may be that reliability professionals have answers to the questions just posed or ways around the obstacles just raised. The point, though, is that we found no one in the mainline professions during our research and interviews even remotely approaching the kind of analysis we are proposing here, though there are reliability professionals ready to analyze issues and other issues this way. That gap needs to be filled, if grid and service reliability is to be enhanced in California, and presumably, elsewhere.

Promoting Careers for Reliability Professionals

The analytic and interpersonal skills needed to undertake "reliability analysis" derive less from the formal professions than from the professionalism of the analyst, and less from specialized training than from careers which span a variety positions and jobs. We now turn to our recommendations on how better to promote these careers in "reliability analysis" and management.

True believers in the logic of engineering and economics can be forgiven for not better understanding that the most effective way to improve grid and service reliability is less through improved designs than through more wide-ranging careers of those professionals doing the designing and implementation. This is a well-known insight not willingly acknowledged by those who believe in structural rather than behavioral approaches to critical infrastructures.

As we have seen in this report, cross-training, rising through the ranks, and shared professional experiences are crucial to ensuring reliability, and even more so networked reliability. Careers built this way allow the professional to undergo the different levels and phases critical to reliability management—field, headquarters; operations, management; emergencies, bad days, normal days; and so on—and we know from reliability research that such experience is key to realizing high reliability in the critical infrastructure. Further, such experience plays an instrumental part in helping to develop elements of a common professional perspective—a frame of reference which facilitates understanding across organizational boundaries and contrasting interests.

How do you build such careers? There are, of course, no formulae. But two recommendations are, as they say, no-brainers.

- *First, restore the financial solvency of PG&E and SCE.* We can think of no other institutional priority as important to ensuring networked reliability than the recovery of these two high-reliability organizations. Their control rooms are the storehouse of multi-level, multi-phased experience essential to the sound operation of a grid increasingly subject to volatile performance conditions. A point made several times in the report deserves repeating here. The real experiment in the California electricity restructuring hasn't been deregulation, but rather the volatile future created by undermining the core competencies of the state's two major reliability institutions in the electricity sector, PG&E and SCE.

- *Second, undertake new multi-disciplinary training immediately.* One significant change that restructuring has accelerated throughout the California electricity system is the increased importance of understanding the basics of the behavior of markets and market fallbacks. Yet we observed significant gaps in understanding the economics of the emerging and morphing power markets on the part of many engineers and operators in the ISO and elsewhere.

To be sure, market analysts and software designers who write market rules into their software are available and we frequently observed them coaching dispatchers about why market economic logic would lead a particular generator to behave in an otherwise strategic way. But this is not a substitute for understanding on the part of dispatchers themselves of what is likely going on from a market or fallback strategy perspective. Further, there has been a sharp rise in the number of individuals with MBA's and economics degrees into the power system with little or no appreciation for the power engineering side of the process. Unless steps are taken this disciplinary gap could widen within a variety of organizations, not only the ISO, with harmful effects.

Accordingly, more time and resources must be devoted to supplementing the basic disciplinary training of HRN and RRN personnel to close the gap between the two disciplines of economics and engineering. This could be facilitated by a series of multi-disciplinary seminars made available to personnel throughout the organizations of the HRN and RRN. These seminars could be funded by fees paid by these organizations.

Improving Recruitment

As we have noted throughout our report, informal networks and information sharing among the various organizations comprising what we have been calling the high reliability network, regulatory reliability network and the reliability task environment are crucial resources in coping with volatility and developing options and strategies with which to better cope and adapt.

Yet over time, as the organizations develop their own separate recruitment and training streams, the basis for common communication and understanding may erode. To prevent this and in the process encourage recruitment into the cadre of what we have been calling reliability professionals, we recommend a program of personnel exchanges among organizations placed throughout the HRN, RRN and RTE. This recommendation has been received with special interest in our presentations to CPUC and ISO staffers. The exchanges would be accomplished by short-term individual postings (six months to a year), in particular between departments of the ISO and distribution units in PG&E and SCE, large generators, and regulatory organizations such as the CPUC, CEC and WSCC. While some generators may worry about proprietary information which could be gained at the ISO or conflicts with market strategies based on withholding information, these would seem to be manageable issues and are, we believe, outweighed by the major advantages of providing transparency to the mutual benefit of all the participants who have a stake in service and grid reliability after Enron *et al.*

Creation of the California Electricity Reliability Network

Who would coordinate such personnel exchanges? Who for that matter would support and lobby for the kinds of analytic capacity building, organizational solvency and career paths outlined in the preceding recommendations? The middle ground needs its own professional organization, as that organization would work at the scope and knowledge base currently unoccupied by any of the formal disciplines crucial for operations in the reliability space .

To that end, we propose the creation of a state-wide Electricity Reliability Network. It would, consist of reliability specialists, practitioners and regulators both to advise the Governor's Office and the State Assembly on proposed policy and management changes affecting the California electricity system as well as to support organizations in the RRN in fulfilling their partial reliability mandates. The Network would be the state's professional association that spoke on behalf of reliability for the California electricity system as a whole. In this way, the Network would lay claim to represent the middle ground niche of reliability professionals who know reliability matters.

The purpose of the Network would be to undertake "reliability analysis" and promote career development and recruitment in ways that bring the results of these efforts prominently to bear on public debate about proposed changes in the future to the California electricity system. Clearly a review of proposals not just for residential metering, but also for forward contracts and distributed generation systems cry out for the kind of "reliability analysis" proposed in this report, as they have substantial implications for system volatility, network options and system operator adaptability. The Network should also sponsor research on all multi-disciplinary aspects of service and grid reliability: engineering, economic, organizational and managerial. It should offer advisory opinions based on this research. It should also issue an annual report on the state of the grid and encourage long-term planning and investment in reliability-enhancing technology and programs by other agencies. It should function above all as an active public advocate for the professionalism and institutional importance of the reliability perspective.

We recommend that the first research assignment of the Network be to rectify the poverty of economic theory underlying deregulation of critical infrastructures, such as electricity. This might sound just like a recommendation academics would make, but it is, we truly believe, the most practical one of this report.

Economic theory is too important to be left to the economists alone. Economists are surely right in pointing out that many of the interventions that have worked to ensure grid and service reliability, such as the price caps, entail efficiency losses. "Reliability analysis", however, interprets these losses differently. The deadweight loss resulting from price caps is one measure of how much reliability matters to us as a society, i.e., reliability must be at least as equal to the efficiency loss. In the same way, the difference between the inelastic demand curve for electricity we now face and the more elastic demand curve we could but do not have with real-time residential metering, must, from the perspective of "reliability analysis", measure the opportunity costs we place on the level of reliability we now have but perceive we would lose if we had the metering. Indeed, one could make very similar arguments that the inelastic demand curve incorporates the premium or overhead we place on reliability, or that this curve represents

the value added by reliability over what would have been the case with a more elastic demand curve, or that the inelastic demand curve reflects the shadow prices with which we hold reliability in a non-competitive market situation.

Economists are also quite right in pointing out that what were fixed costs in the short run become variable over the longer run, i.e., there is no such thing as reliability's non-fungibility over the long haul. "Reliability analysis", however, interprets the overwhelming evidence of electricity's non-fungibility in real time as evidence that the short-term is a persisting feature of electricity as an always-on service and as a critical infrastructure. In the same way, if economists are right that fixed costs should not matter in marginal analysis, then reliability professionals must conclude that marginal analysis is not appropriate for making decisions about the non-fungibility of electricity in real time.

Economists are also surely correct in pointing out that typically such infrastructure as electricity has had a low discount rate spread over a very long term in which that infrastructure produces streams of benefits and costs. "Reliability analysis", however, underscores the fact that reliability is treated as if it had an exceptionally high discount rate, where the present without reliability matters more than the future with it. Cost-benefit analysis assumes that per capita incomes are going up over time, when in fact what is at issue is that per capita incomes would assuredly go down if reliability did not matter as much as it does in the present.

What is at issue here, in other words, is not that we need less economic analysis in the electricity sector. The more analysis, the better. What we do need less of, however, is interpreting the results of this analysis outside a "reliability analysis" framework by economists and other specialists who are not even trained as reliability professionals. What we have been calling "reliability analysis" must precede and serve as the context for economic analysis because, while markets coordinate through the price signal, HRN operators and wraparounds are responsible for ensuring that the premises and assumptions (e.g., software embedded rules) are in place so that coordination can take place. The effort to ensure market premises that are in place under shifting conditions of system volatility and network options variety cannot itself be coordinated by the very same market. Markets that operate within a context that takes grid and service reliability seriously are the only reason why we could ever expect grid and service reliability to be enhanced through those markets. It is hoped that the proposed Network would develop and make this point emphatically, particularly with respect to developing standards for market reliability analogous to those WSCC criteria that already exist for grid and service reliability.

Conclusion

This report has been all but silent on the implications of our research findings for the regulators and the general public, or what this report, which uses many acronyms, has been calling the regulatory reliability network (RRN) and the reliability task environment (RTE). This silence has been deliberate, as the report focuses on the organizations and units that actually provide the electricity, or what we have been calling the high reliability network. Indeed, if we had more funds, researchers and time, we would not only have studied the regulators but also interviewed representatives from the municipal utilities and irrigation district generators in order to update the picture of the HRN.

Several implications for the RRN, however, follow directly from the preceding analysis in this report and it seems fitting to draw them out in the report's conclusion.

First and foremost, undertake “reliability analyses” of proposed regulatory interventions and ensure each passes the reliability-matters test. If there are trade-offs between interventions that affect volatility, options and adaptability, make them explicit, so that all can understand why inducing a negative change in one must be more than compensated by inducing positive changes in the others. It may be that many in the RRN have actually produced a set of rules and interventions that, in hindsight, are successful in meeting the reliability-matters test, and it is this set that provides useful lessons for the future.

Second, focus on the longer-term issues where reliability is fungible and in trade-offs with other factors. It is clear to us that the HRN will be preoccupied with short-run, indeed real-time, issues for the foreseeable future. Similarly, consumers and others in the RTE understandably have shorter time horizons. Thus, it is up to the regulators to take the longer-term perspective in the California electricity system, if that perspective is to count. Indeed, if major redesign of the California electricity system is being contemplated, we would recommend rethinking the role of regulators in the system to ensure regulation's coherent shift to other than short-term considerations. Our analysis suggests that strategies such as regional and statewide load limits might reduce overall volatility and better ensure reliability over the longer haul.

Third, install real-time grid status indicators among all HRN and RRN participants. One major finding in the research is the enormous operational virtuosity required to maintain both service and grid reliability under the restructured electricity system. But an equally striking finding is how little the challenge of grid reliability is factored into decisions taken by actors outside the ISO, and how poorly market pricing reflects the "overhead" costs of this reliability.

One way to begin to address this basic market imbalance is to provide information to as wide a set of participants as possible on the state of the grid. In our interviews in power generation facilities, large wholesale power customers and even in the dispatch control centers of distribution units, the absence of any real-time indications of the area control error (ACE) was quite conspicuous. When asked about this, many operators and supervisors simply noted that given their operational responsibilities they did not need to follow the ACE.

It would seem that a basic initial step toward having all participants internalize the issue of grid reliability is to require all large generators and distributors of power in the California market to have a real-time display of the ACE. Further, we think it worthwhile to attempt to develop a set of additional key indicators of the state of the grid such as major pathways in and out of service, Path-15 conditions (including temperature) and current regulatory contingencies applied to the operation of the grid (such as timed CPS2 violations) that would be widely available on a readily accessible web page. If the proposed California Reliability Network does not undertake the task, we would hope key regulators in the state could agreed to do so among themselves.

Finally, like all policy analysts, we have to ask: What else is missing from these recommendations? For all our criticisms of restructuring, the close reader will have noticed that we do not recommend anything like the re-regulation of the electricity sector. We know of no

organizational reforms that would put Humpty-Dumpty back together without increasing substantial system volatility and other real-time reliability problems in the process. We live in a world of networked reliability where market entities and their fallbacks play an important role. Not only important, but one whose behavior is volatile for the foreseeable future, as every casual reader of the headlines around Enron *et al* knows. In the meantime, no further major system redesigns of the high reliability network should occur without all participants confronting the reliability challenges raised in this report.

A

SHORT HISTORY AND INTRODUCTION TO THE CALIFORNIA ELECTRICITY RESTRUCTURING, CRISIS AND SYSTEM

California's Pre-Restructuring Electricity Sector

As with the rest of the US electricity industry, California's grew explosively during the first half of the twentieth century. From the 1930s until the 1990s, the state's electricity industry was dominated by three large-scale, vertically integrated, investor-owned utilities (IOUs). Pacific Gas and Electric (*PG&E*); Southern California Edison (SCE); and San Diego Gas and Electric (SDGE) owned and operated the three elements of the state's electricity infrastructure: generation, transmission and distribution. Next to these IOUs, various smaller regions in California were served by state or municipal integrated utilities.¹

Both the state through the California Public Utility Commission (CPUC) and the federal government through the Federal Energy Regulatory Commission (FERC) regulated the electricity industries' activities to ensure that electricity was provided at "reasonable rates". The CPUC could set and change retail rates on demonstrated costs, control the utilities' service reliability, service to customers, and control facility siting for new generation (Kahn & Lynch, 2000:15; California State Auditor, 2001:9). FERC, for its part, regarded the increasingly integrated electricity grids as "a natural monopoly in interstate commerce subject to federal regulation" (Energy Information Administration, 1996:106) and held regulatory powers over the interstate wholesale electricity trade.

During the 1950s and 1960s, California's growing economy and population contributed to a flourishing electricity industry.² Rising electricity demand, technological innovations and economies of scale were implemented, leading the utilities to invest in large, centralized power plants. In these years the electricity grids of the utilities were interconnected and generating capacity was greatly expanded. A massive cascading blackout in 1965 in New York brought home the vulnerability of these interconnected systems. This triggered the formation of the voluntary North American Electric Reliability Council (NERC) and the Western Systems Coordination Council (WSCC) to provide oversight and standards for the reliability of electricity grids on a national and regional basis respectively.

1 They represent over 35 municipal and co-operative utilities, and three small investor-owned utilities. Among them: Sacramento Municipal Utility Department (SMUD), Los Angeles Department of Water and Power (LADWP), and Bear Valley Electric Service. These utilities were also regulated by the CPUC, and together accounted for about 30% of the retail electricity sales in California (Joskow, 2000a:117; Congressional Budget Office, 2001:25).

2 California Public Utilities Commission (1993), Chapter III, pp. 17-26.

The economic recession of the 1970s hit the California's electricity industry particularly hard. Rising fossil-fuel prices in the 1970s and early 1980s led to subsequent electricity price increases. The sector also suffered high cost overruns for new investments—especially new nuclear power plants³—as well as high inflation and interest-rates. California nominal electricity prices more than doubled from 1965 to 1981 from 2 cents per kilowatt-hour to over 5 cents in 1981 (California Public Utilities Commission, 1993:31). By 1980, the nominal electricity prices of the state's three utilities were well above the national average, as can be seen in Figure A-1.

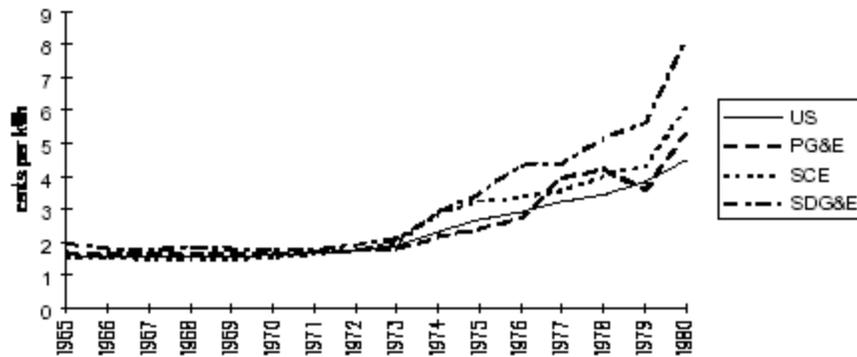


Figure A-1
Average Electricity Rates 1965-1980 (in Nominal Dollars, cents per kWh)
Source: California Public Utilities Commission (1993), pp. 32, Figure IV-4

As a result of such factors, California's utilities imported cheaper power more and more through its interconnections (interties) with other states, with a concomitant increase in FERC's authority over wholesale interstate transfers. In the north, the state of Washington and its Bonneville Power Administration (BPA) had constructed large dams that seemed to provide an abundant amount of hydropower, which could help California cope with peak summer demand. During the winter, California's reduced demand for electricity provided the north with residual power to cope with its own peak winter demand.

Higher electricity costs caused mounting financial problems for the utilities. By 1979 the utilities were in poor financial shape due to these rising costs and debts, which could not be recouped through higher retail rates that were closely regulated by the CPUC. To add to the problems, electricity demand stabilized and even declined, leading to a comparatively abundant supply of generation. The specific problems that plagued the Californian electricity industry caused the CPUC to radically change its electricity policy in at least three ways:

- Increasingly, the CPUC began managing the utilities' investments and operational decisions, further adding to an already complex regulatory setup.
- California undertook "integrated resource planning" through which state energy planners of the newly created California Energy Commission (CEC) and regulators of the CPUC were to balance future supply and demand, building new power plants when needed but investing in

³ For example, the costs of the Diablo Canyon nuclear power plant were originally estimated at \$500 million, but ultimately exceeded \$5 billion (California Public Utilities Commission, 1993:81).

conservation and energy efficiency to minimize the need for costly new plants (Kahn & Lynch, 2000:35).

- The CEC and CPUC tried to remedy the electricity industry's reliance on fossil fuels by vigorously implementing the Public Utility Regulatory Policy Act (PURPA) of 1978 and actively promoting fuel and resource diversity in the industry. Importantly, PURPA enabled non-utility investment in electricity generation, focusing on renewable resources⁴ and the use of co-generation⁵. CPUC and FERC shared responsibility in regulating the non-utilities, the so-called Qualified Facilities (QFs). The CPUC required utilities to buy relatively more expensive "green power" from the non-utility generator at regulated rates, which was to guarantee the non-utilities a high and stable profit-margin in selling their electricity to the utilities (Energy Information Administration, 1993:10; Kahn & Lynch, 2000:35/41; World Bank, 2001:3).

From the middle 1980s, general economic conditions steadily improved in the state. Stable and lower fossil-fuel prices allowed the utilities to recover from the 1970s recession, with the CPUC concentrating on stabilizing the electricity industry and its electricity retail prices as can be seen in the trend in Figure A-2.

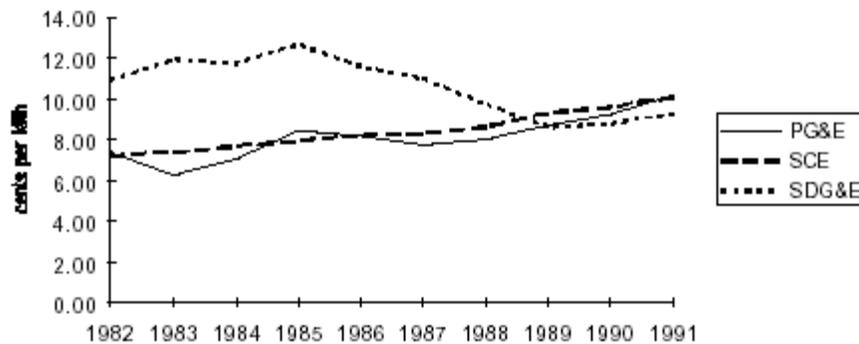


Figure A-2
Average Electricity Rates 1982-1991 (in Nominal Dollars, cents per kWh)
Source: California Public Utilities Commission (1993), pp. 66, Figure V-6

A specific CPUC measure that helped utilities regain their financial health was its policy aimed at delaying power plant investments. Instead of building new power plants to meet rising electricity demand, utilities were encouraged to invest in energy savings⁶ and consumer conservation programs, significantly reducing the need for major investments in California power plant construction in the 1980s. However, electricity prices remained among the highest in the U.S.

The need for additional generation resources returned during the late 1980s with increasing electricity demand and renewed economic growth. The explosive growth of non-utility (QF)

⁴ These include wind, solar, biomass, small hydroelectric generation or efficient fossil-fuel technologies.

⁵ The production of electricity here is the by-product of a manufacturing process through the use of heat or steam.

⁶ CPUC measures resulted in a 250% increase in conservation investments between 1980 and 1985, saving an estimated 12 billion kilowatt-hours (California Public Utilities Commission, 1993:71/79).

generation in California could cope with these demand increases, however. Most non-utility generators were selling their excess electricity directly to electric utilities, which proved to be highly profitable. By 1991, 87% of the total U.S. non-utility production of renewable sources was produced in California, amounting to 53 billion kilowatt-hours and accounting for 34% of the total electricity production in California (Energy Information Administration, 1996:10-13).⁷ Together with increasing electricity imports from neighboring states, these additions were enough to meet California's growing electricity demand.

As early as the middle 1980s, the CPUC expressed doubts about the viability of its regulatory structure. The increasing time, complexity, and administrative costs of the type of regulation employed caused the CPUC to slowly shift its regulatory regime from rate of return regulation to performance regulation, relying less on governmental management and more on market incentives. The CPUC actively began encouraging competition in the electricity industry, using new regulatory frameworks developed for the liberalized and deregulated telecommunications and natural gas industries. In the early 1990s the CPUC's commitment to integrated resource planning, energy efficiency and fuel diversity was replaced by the use of a competitive bidding process to promote new electricity generation construction.⁸

The combined developments—California's increased reliance on electricity imports from other states and electricity generation by non-utilities and active promotion of market incentives by the CPUC—created a relatively competitive and high-volume wholesale market. It is therefore not surprising that California was among the first states to respond to growing pressures to increase competition in the electricity industry. Following the overall national trend of strong customer dissatisfaction and spurred by California's high electricity prices in the beginning of the 1990s, the CPUC became interested in and started promoting further competition in electricity generation as a method to reduce electricity production costs. In September 1992 the Commission announced its intent to “examine the conditions the electric industry currently confronts...and explore alternatives to the regulatory approach” in California (California Public Utilities Commission, 1993:87).

Restructuring the Industry (1994-1998)

California's electricity industry slowly evolved into a hybrid structure in which regulation and competition played important roles. California's restructuring debate centered around the issue of if and how deregulation and increased competition would substantially lower the relatively high electricity prices within the state. The result of the restructuring debate was that California entirely restructured its electricity industry and market. The core of the restructured (popularly called “deregulated”) wholesale industry was formed around a new electricity market and a new industry structure in which competition and market incentives would take over regulatory and command-and-control mechanisms.

⁷ In 1991, the percentage of total electricity requirements purchased from non-utility generators by the three large IOUs reached 32% for SCE, 24% for PG&E and 6% SDGE (California Public Utilities Commission, 1993:58, note 70).

⁸ Nevertheless, until the electricity industry restructuring plans were implemented, no significant amount of new generation capacity was built.

To further the transformation, the CPUC moved swiftly after the publication of the Federal Energy Policy Act (EPACT) in 1992, ordering an examination “of the comprehensive set of regulatory programs in order to explore alternatives to the regulatory approach based on conditions and trends identified in a previous decision” (Energy Information Administration, 1996:67). The result was a staff report known as the “Yellow Book” in February 1993.⁹ In this paper, the CPUC described the existing regulatory system as “fragmented, outdated, arcane and unjustifiably complex” (quoted in World Bank, 2001:3). Two points are important to note here for the purposes of this report. First, the CPUC sought to redefine the prevailing regulatory compact¹⁰ and a fundamental restructuring of the electricity industry. We thus see redefinition of electricity reliability at the heart of and fundamentally fueled by the restructuring process. Second, virtually all stakeholders—consumers’ organizations, large industrial consumers, utilities and non-utility electricity producers, among many others in the RTE—within California’s electricity system received the Yellow Book favorably.

Consequently, from early on California’s debate focused primarily on how to restructure, not if restructuring was necessary (Hawk, 1999:71). The steamroller of restructuring at that time is nicely captured by interviews with two senior staff members of the CEC and CPUC respectively:

There was definitely a feeling that if you were not supporting restructuring, you were not listened to. People did not want to hear expressions of concern about restructuring, or thinking it would not work. That was not a welcomed point of view at that time.

We went into this with a consensus model... There was a kind of groupthink and anyone with a different view was really admonished. . . Even environmentalist and consumer groups went along with the consensus view—groupthink, not making hard decisions, but trying to make everybody happy.

The CPUC began investigative proceedings to consider its proposed restructuring policies in April 1994. These initiatives, known as the “Blue Book” proposals, outlined a strategy and regulatory structure to replace traditional cost-of-service regulatory framework with alternatives that focused on utility performance and, where possible, the discipline of the market.

It did not take long before dissension arose leading to a “polarization among stakeholders regarding whether restructuring and regulatory reform should occur at the wholesale versus the retail level and how centralized decision making should be” (Hawk, 1999:75-76). First, there was the issue of whether California’s electricity industry would be fully competitive or whether competition would remain limited to the wholesale market. Second, there was the question of the electricity structure and operations. Two structures of California’s electricity sector were proposed: the PoolCo model and the Direct Access Model.

⁹ California Public Utility Commission (1993), *California’s Electric Services Industry: Perspectives on the Past, Strategies for the Future*, San Francisco CA.

¹⁰ The regulatory compact of the IOUs ensured that IOUs obtained “an exclusive franchise to provide service to retail customers within its service territory and the privilege of regulated prices to earn it a fair rate of return on its investment” (Energy Information Administration, 1993:3). See also California Public Utilities Commission (1993:9).

The PoolCo model embraced a single institution that would operate the transmission grid and hold a daily auction to determine generation schedules with the joint goals of minimizing short-term operating costs of supplying California with electricity, while maintaining a reliable transmission system operation. Its approach to scheduling was similar to that traditionally used by the IOUs in that it built a “least-cost” schedule from cost price information that respected transmission system constraints.

The Direct Access model envisioned one transmission system operator, whose principal goal would be to maintain reliable transmission system operation while limiting its involvement in the electricity commodity market. Multiple market mechanisms would be expected to emerge to match electricity suppliers’ bids and customers’ demands and determine generation schedules and prices. Design priority was based on decentralized decision making in generation scheduling and dispatching, with the expectation that competition between multiple mechanisms would lower longer-term and eventually short-term generation prices to consumers (Hawk, 1999).

The proposed structures differed on fundamental points. Chief among them was the role of the system operator in creating and managing organized public electricity markets. The PoolCo model more or less resembled the so-called power pool structure that was already functioning in parts of the U.S. and Europe.¹¹ The system operator would be granted with large authorities and consequently would have a large influence on the electricity market. The Direct Access model, on the other hand, consisted of a radically new market structure that would allow far more freedom for electricity trading in the form of bilateral trading outside of the power market.

The proposals also differed in terms of institutional, operational and financial arrangements made for scheduling generation and the actual operation of the transmission system. The PoolCo model was developed and preferred by two IOUs: SCE and SDGE. Other supporters for the model were the CEC and several governmental organizations. The Direct Access model was developed by PG&E and supported by energy producers, power marketers and most large commercial and industrial electricity consumers (Hawk, 1999), who criticized the PoolCo model as a move toward “Soviet-style” central planning (Bushnell & Oren, 1997:237).

In May 1995, the CPUC issued a proposed decision consisting of two proposals: a majority Proposed Policy Decision (PPD) based on the PoolCo model, backed by three Commissioners and a minority Alternate Proposal based on the Direct Access model issued by Commissioner J. Knight Jr. Subsequent Restructuring Proceedings of CPUC sought the opinions of stakeholders on both proposed models and showed increasing tensions between stakeholders involved in California’s restructuring process.

One of the major issues concerned the ability of the utilities to use long-term contracts. One regulatory official described the ruling fears among new non-utility generators, power marketers and regulators alike that “it was thought that the utilities would strike private backroom deals with the new merchant generators”, thereby killing every attempt at creating a competitive

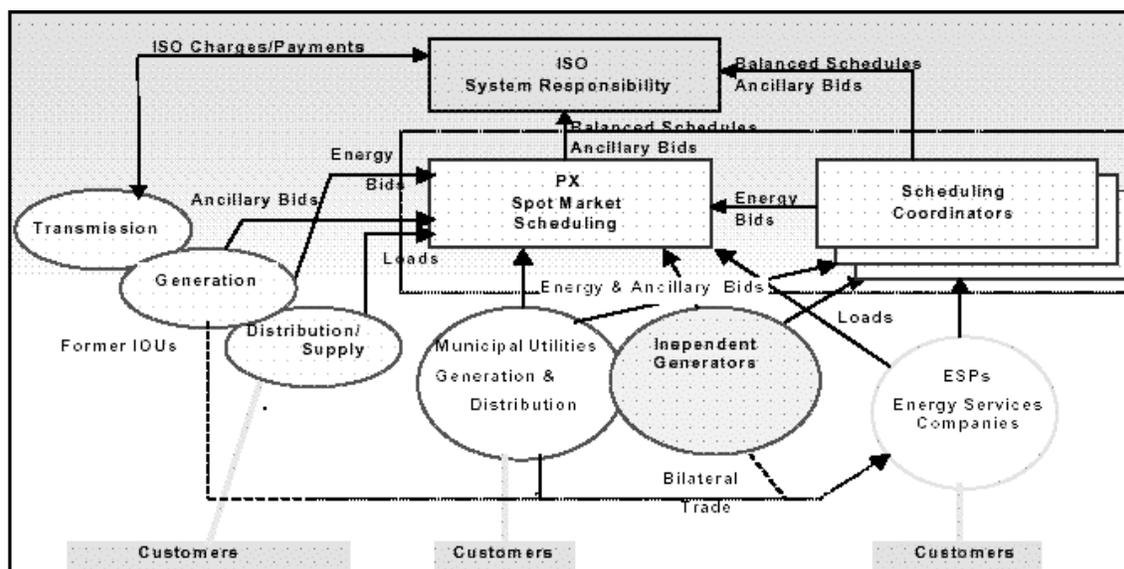
¹¹ Examples included the Pennsylvania, New Jersey, Maryland pool (PJM) and the previously restructured English and Scandinavian power pools. Several of the then-sitting CPUC commissioners were greatly influenced, by their visit to England and Wales in early 1994 to study the competitive electricity system created there in 1990 (Joskow, 2000b:42, note 34).

wholesale market and competition between electricity generators. The utilities on the other hand were adamant about compensation for their stranded investments in “green power” contracts and expensive power plants.

In an attempt to break the stalemate in California’s restructuring process, the governor’s office initiated discussions and negotiations between the parties who represented the two poles¹² in the debate and a solution was reached in a subsequent Memorandum of Understanding (MOU). The IOUs were allowed to recover their “stranded costs” (i.e., anticipated above-market costs of previous investments) associated with operating the two high-cost nuclear power plants and the state-mandated purchases of power from the relatively high-cost renewable power generators through a “competitive transition charge” on consumers’ electricity bills. In return, the utilities agreed to the establishment of a hybrid market structure in the form of a Direct Access-type model that would allow both bilateral and market deals and give retail customers the choice to obtain electricity from any utility or other Energy Service Provider (ESP) (Bushnell & Oren, 1997). However, the utilities would initially be forbidden to enter into long-term bilateral contracts and would be obliged to procure all their electricity through the newly established electricity market. This opened the way for CPUC’s final decision in December 1995, which laid out a set of policies to create a fully competitive electricity market and to guide the utilities in restructuring their operations (California State Auditor, 2001:8). This would be supported by the introduction of full retail competition and the vertical unbundling of the electricity industry to enable competition (Joskow, 2000a:139).

The New Structure

The restructuring program was carried out from early 1996 and created the new electricity industry structure schematically outlined in Figure A-3.



¹² These parties included SCE, two organizations representing large electricity consumers and an association representing non-utility generators (Hawk, 1999).

Figure A-3
California's New Electricity Market Structure Established by AB-1890
Source: Marnay (*et al*) (2001), p. 6, Figure 4

California's electricity sector was to be composed of a novel mixture of private and public organizations that were involved in the provision of electricity. Not all the municipal utilities in California were incorporated in this new system and some have retained control of their power plants and transmission systems (Kahn & Lynch, 2000:20).

California's restructuring was to mean a competitive and fully unbundled electricity production chain. Under the CPUC's final decision, the IOUs were "encouraged" to divest most of their in-state fossil-fuel generation and sell it to new "merchant" power producers, also known as Independent Power Producers (IPPs) or Independent Generators.¹³ Electricity generation would become the business of competitive, independent electricity generators. These private generators would consist of the IOUs former (fossil fuel) generating plants that had largely been bought by new merchant generators and the co-generation and renewable non-utilities. Operational control of the utility-owned high-voltage transmission grid was transferred from the IOUs to the California Independent System Operator (ISO), a hybrid not-for-profit public-benefit corporation.¹⁴ Nothing quite exemplifies the experimental thrust of the California restructuring exercise than this novel, hybrid organizational form for its independent system operator.

The ISO was to be responsible for balancing load and generation throughout the grid so as to guarantee a reliable and high-quality provision of electricity.¹⁵ To do this, the ISO would have to make (near) real-time adjustments in amount of electricity fed into the grid by ordering electricity generators to increase or decrease their electricity output. The restructuring envisioned the ISO performing this function through the use of a Real-Time Imbalance Market, an Ancillary Services Market and a Congestion Market. These markets would consist of various services to attain the necessary electricity and "reserves" to reliably operate the high-voltage electricity grid based on how quickly they can be made available if needed.

The ancillary services market consisted of the following types of electricity generation:

- Regulation: generation that was already up and running and could be increased or decreased instantly to keep supply and demand in balance;
- Spinning Reserve: generation that was running, with additional capacity, which could be dispatched within minutes.
- Non-spinning Reserves: generation that was not running, but could be brought up to speed within ten minutes.
- Replacement Reserves: generation that could begin contributing to the grid within an hour.

¹³ CPUC Decision (D.) 95-12-059. Formally, the IOUs were required to divest at least 50% of their fossil-fuel plants. Hydroelectric and nuclear power plants were allowed to remain under IOU ownership. A senior CEC staff member stated: "But they did even do better than that. They got rid of virtually all of their fossil-fuel generation."

¹⁴ See the ISO website (as of the time of writing): <http://www.caiso.com/PowerCentral/>

¹⁵ Before restructuring, each vertically integrated investor-owned utility performed the grid management functions for their own specific geographical area. In other areas, utilities centralized these functions in a power pool.

Two other types of ancillary services—voltage support and black start—were procured by the ISO through bilateral, long-term “Reliability Must Run” (RMR) contracts to ensure the ISO’s control of the grid. These contracts were made with power plants that the ISO judged essential to the grid based on their relative output and location within the grid in accordance with WSCC reliability criteria. The ISO directly controlled these units to operate the grid and maintain a balanced load. Generators could submit bids for these markets. RMR units, however, were to be used for local, rather than grid-wide, support.

The ancillary services market was thought to accommodate about 3 to 5% of the total electricity needs of the state and provide the ISO with the necessary reserves to reliably operate the high-voltage electricity grid (California State Auditor, 2001:16). When ISO needed more energy than what was bid in the ancillary services and real-time imbalance markets, the ISO could make “out of market” (OOM) purchases, typically from generators in other states or municipal utilities.

To ensure the ISO would be provided with enough operating reserves—the capacity needed to keep the grid reliable—a system of warnings and emergency alerts was devised. When forecasted reserves for the next day would fall below 7%, the ISO would issue an Alert, and generators would be asked to increase their power bids into the market.¹⁶ When forecasted reserves for the day fell below 7%, the ISO would issue a Warning, and the ISO would start buying these supplies directly. When actual reserves would fall below 7%, 5% and 1.5% respectively, the ISO would issue a Stage 1 emergency (public appeals and other measures to increase supply and decrease demand), Stage 2 emergency (interruptible customers curtailed), and a Stage 3 emergency under which firm customers would be blacked out to keep the system from crashing (Kahn & Lynch, 2000:17-18).

Under restructuring, the IOUs would remain the owners of the transmission network grid (and their distribution grid as well) in their service area and would be transformed into Utility Distribution Companies (UDC’s) (California State Auditor, 2001; World Bank, 2001:3). Retail rates were to be frozen for four years (until March 31, 2002) or until the utilities’ stranded costs were recovered in the expectation that restructuring and the introduction of competition would significantly lower electricity prices.¹⁷

Last but certainly not least, a new intermediary group was to enter the electricity industry: the electricity schedulers, also known as energy brokers or, more commonly, scheduling coordinators (SCs). The task of the SCs was to match electricity supply and demand in a market setting. Matched blocks of supply and demand were to be submitted to the grid operator, ISO, which calculated whether or not the transactions met the capabilities of the high-voltage transmission grid. Another new institution, the California Power Exchange (PX), itself a hybridized private non-profit organization, was to function as California’s main SC and the primary wholesale electricity market where supply and demand would meet. The PX would consist of markets in which electricity could be traded at various prices. Of these, the day-ahead

¹⁶ The ISO could also issue “No Touch” and “Power Watch” messages to the industry to warn of impending problems and be able to acquire the amount of electricity needed to maintain grid reliability.

¹⁷ The California State Auditor mentioned an expected overall reduction in electricity rates of *at least* 20 percent by April 1, 2002 (California State Auditor, 2001:9).

market was to be the largest and most important one with an anticipated volume of 90% of all trades (California State Auditor, 2001). Prices in these markets were based on an hourly basis:

In this market, buyers requested the amount of electricity they anticipated needing for each hour of the next day and stipulated the prices they were willing to pay. At the same time, sellers stated the amount of energy they could produce and the prices they required for each of those hours. Once the PX had received all of the demand and supply bids, it matched them, thereby acting and functioning as a large scheduling coordinator. The highest-priced supply bid necessary for meeting demand during any given hour would set the single market-clearing price to be paid by all buyers to all sellers for energy purchased for that hour” (California State Auditor, 2001:9).

In contrast to the non-utilities and merchant power plants—who could purchase their electricity through the PX or through bilateral contractual agreements with electricity generators or the more than 40 other scheduling coordinators—the state’s IOUs were required to sell and purchase all of their power through the PX until March 2002 or until the CPUC ruled that they had recovered their stranded costs.¹⁸

In creating the PX and ISO, restructuring essentially federalized the California grid. The responsibility of electricity generation pricing and some authority over grid management—for example reviewing and approving tariffs and interstate access—was removed from the state level (CPUC) to the federal level (FERC). At the state level, a new regulatory agency was created under the jurisdiction of the state governor to monitor the newly created institutions and the new market: the Electricity Oversight Board (EOB).¹⁹

To share the gains of restructuring with consumers, the CPUC set electricity retail prices at 90% of the 1996 retail prices. California’s State Legislature passed AB 1890 in September 1996, codifying many of the CPUC’s policies and recommendations for California’s electricity restructuring. The key milestones of the restructuring exercise described above are summarized in Table A-1.

Table A-1
A Chronological Outline of California’s Electricity Restructuring Process

Event	Date
CPUC Strategic Planning (Yellow Book issued)	February 3, 1993
CPUC Restructuring Proceedings opened (Blue Book issued)	April 20, 1994

¹⁸ From June 1999, the PX also offered a block-forward market, in which market participants could make longer-term deals for electricity at set prices. The CPUC had to approve the amount of energy that investor-owned utilities wished to purchase using this arrangement. From December 2000, FERC lifted the state’s requirement that the IOU’s were obligated to buy and sell power through the PX (California State Auditor, 2001:9).

¹⁹ The EOB—consisting of state legislators and governors appointees—appointed the governing boards of the PX and ISO. Other tasks of the EOB included overseeing, the operation and reliability of electricity transmission systems, the operation, efficiency, and competitiveness of markets for bulk-energy transmission, and the activities of the ISO and Power Exchange. Furthermore, the EOB would be reviewing market and reliability rules, maintenance, repair, and replacement standards and transmission and grid plans. At the time of writing, the role of the EOB had been minimal in the RRN.

CPUC Proposed Design	May 24, 1995
Market Structure Memorandum of Understanding (MOU) signed	September 14, 1995
CPUC Final Decision	December 20, 1995
State Law AB 1890	September, 1996
California's Restructured Electricity Market Opens	March 31, 1998

Source: Hawk (1999), pp.71

It must finally be noted that California's economy and subsequent electricity demand grew at a dramatic rate during the years of restructuring. Peak demand increased 18% between 1993 and 1998. During those same years, only 0.1% of generating capacity was added, causing California's reserve margins (i.e. excess generation capacity) to fall from 13% to 4% (Faruqui *et al*, 2001: 38). Throughout 1997, the various entities prepared themselves for the new market structure. The IOU's divested most of their fossil-fuel power plants, reducing their percentage share in electricity generation from 81 to 45.8% in 1999.²⁰ At the same time, the ISO and the PX were built from scratch to take over control of California's electricity industry operations.²¹

The First Years of Operation (1998-2000)

During its first two years of operations—apart from some start-up problems—the California electricity market seemed to be working mostly as designed and expected.

Although originally planned for January 1, 1998, California's new electricity structure started functioning on March 31, 1998. This delay was caused by problems in the start-up of the ISO, the PX and their respective markets. The markets were developed and created during little more than nine months and not all the bugs were solved by the time operations began (Joskow, 2001a:18). During the first months of operation, the electricity market as a whole was hampered by a number of design problems, causing dysfunctions in some parts of the electricity market.²² For example, major flaws were identified in the protocols for planning and investment in transmission and in the markets. Important operational problems arose with respect to “overgeneration”, non-compliance, and the ancillary services market.

By the spring of 1998, record rainfall in the previous winter had caused an abundance of relatively cheap hydropower in northern California. This relatively low-cost electricity pushed the more expensive gas-fired power electricity in the entire state out of the market and caused regional reserve shortages in the south of California. This necessitated the ISO to order some

²⁰ Energy Information Administration, State Electricity Profiles—California, to be found at the time of writing at the website: www.eia.doe.gov. According to the CPUC, the total revenues from these divestitures amounted to some \$1.356 billion for the three IOUs (At the time of writing to be found at the webpage www.energy.ca.gov/electricity/divestiture.html).

²¹ For a more detailed description of California's electricity market, see Joskow (2000); Kahn & Lynch (2000); California State Auditor (2001).

²² See for example: Hawkins, D. (1999), *Reliability through Markets, California Independent System Operator experiences*, presentation at the April 6 American Power Conference, Chicago, to be found at the time of writing at the webpage, <http://www.caiso.com/docs/1999/05/04/199905040846091356.pdf>.

plants online in the south, further adding to the “overgeneration” problem. “Overgeneration” was also caused by non-compliance. Electricity producers found that “uninstructed deviations” from schedules could be profitable, albeit in the process causing problems in balancing load and generation due to these unplanned deviations. Both “overgeneration” and non-compliance caused the ISO to purchase even more energy reserves from the ancillary services market to balance the grid. Furthermore, ISO operators were coping with the new uncertainties of the electricity markets, in which the system’s load and generation in real time were inherently more unpredictable. Consequently, the ISO purchased far more ancillary services than under the old vertically integrated structure (Joskow, 2000a:172).

During the first months, the amount of services bid into the ancillary services market was too low, leading to abnormally high prices for the services and indicating market power problems. For example, in July 1998, the ISO had to buy ancillary services for real-time balancing for as much as \$10,000 per MW (Joskow, 2000a:170). The unusually hot summer of 1998 and the relative scarcity of reserve power in the ancillary services market necessitated the issuing of Stage 1 and 2 emergencies declarations by the ISO, causing some customers interruptions in electricity because of shortages, though nothing like the number of declarations in the California electricity crisis in 2000-2001 (see Table A-2). The overall result was that ancillary services costs in California during 1998 were much higher than anticipated, rising to over 15% of total energy costs during the summer of 1998. By the fall of 1998, the ancillary services market was restructured, and the market seemed to be working properly during 1999, consequently reducing the percentage of ancillary services costs (Joskow, 2000b:102).

Notwithstanding improvements, the ISO and the PX repeatedly warned the CPUC and FERC from August 1998 onwards of serious flaws in California’s rules and market structure, mentioning in particular strategic behavior, the mandated use of short-term markets (in the absence of the utilities being able to make forward contracts to cover their load), and the lack of demand elasticity in the electricity market (Kahn & Lynch, 2000; State Audit Report, 2001).²³ By 1999, additional worries emerged. In March of that year, the PX concluded that during periods of high electricity demand, market power could determine and set wholesale prices²⁴, thereby voicing its concerns about the spot market price volatility. Furthermore, the ISO began to express concerns about the rapid growth in electricity demand, the rapid reduction in reserve margins and the slow pace of new generation investments (Joskow, 2001a:18-20).

To remedy these problems, the ISO and the PX sought to change the markets and their procedures. Within the first two years of operation, the ISO had filed 30 major revisions to its protocols with FERC. Responding to these series of problems and proposed fixes, FERC ordered the ISO to seek to identify and implement fundamental reforms rather than just piecemeal fixes to individual problems as they arose (Joskow, 2001a:18-20).

That said, over the first few years of their operation the PX and ISO market displayed a “normal” pattern in which wholesale energy prices generally followed electricity usage during the day and the seasons. Electricity prices were lower during the night and winter compared to the day and summer, when people’s activities and climate in terms of electricity usage change considerably.

²³ California State Auditor (2001), pp. 43, Table 5.

²⁴ General Accounting Office (2001) and Joskow and Kahn (2001).

Real-time energy prices, although more volatile and peaking in times of greater demand were roughly moving with day-ahead energy prices (Joskow, 2000a: 166). Competitive wholesale market prices for power were reasonably close to pre-restructuring projections, averaging 3 cents per kWh (\$30 per MWh) between April 1998 and April 2000. In March 2000, the CEC published projections for wholesale market prices for 2000 and beyond, which were in the \$30 to \$35 per MWh range (Joskow, 2001a:23-24). Wholesale price rates remained below the frozen retail rates of about \$65 per MWh retail prices, enabling the former IOUs to make good their “stranded costs”. In fact, SDGE was the first of the three former investor-owned utilities to recoup its stranded costs. Its retail prices were thus freed by the CPUC in mid-1999, meaning that the utility would have to compete in price with other electricity retailers (Kahn & Lynch, 2000:17). For consumers the prospect of free retail prices and competition seemed to fulfill the original promises of deregulation and the passing on of lower wholesale prices in retail prices.

In general, restructuring seemed on track. The PX and ISO markets operated as designed and expected (California State Auditor, 2001) and most encountered problems were solved by changing and adding procedures and market rules. Monthly wholesale prices remained relatively stable during 1998 and 1999 and in 2000 prospects for declining overall wholesale prices seemed favorable.

Those prospects were a dreamland.

California’s Electricity Crisis (2000-2001)

As of 2000, California’s electricity grid spanned some 40,000 miles of high-voltage electricity wires that connected the distribution grids with the generation plants, and connected California with other states and its neighboring countries. California numbered about 1,000 generation facilities with roughly 55,000 MW of capacity. On top of that, the state was able to import an additional 8,000 MW of which 4,500 MW could be considered “firm supply contracts” (Kahn & Lynch, 2000:12). California’s market seemed highly competitive with a relatively large amount of new merchant generators as can be seen in Table A-2.²⁵

Table A-2
Merchant Generation in California

Company	Generation capacity	Percentage market-share
AES	4,071 MW	19%
Calpine	871 MW	4%
Duke	2,950 MW	14%
Destec	1,169 MW	6%
Dynegy/NRG/Destec	1,550 MW	7%
Reliant	3,531 MW	17%

²⁵ The new merchant generators had entered the electricity market by the year 2000. Most of them had bought the divested power plants from the IOUs.

Southern	3,065 MW	14%
Others	4,025 MW	19%

Source: Kahn & Lynch (2000), pp. 16

The electricity market had so far seen huge increases in both the amount of electricity traded and the trading volumes. However, all this was about to change in the first year of the new millennium.

As temperatures rose during the spring of 2000, the electricity market started to run into trouble. Both California and the entire western region experienced one of the hottest summers in decades while hydropower reserves in the Northwest were low due to an extremely “dry” winter (California State Auditor, 2001: 59-61; Congressional Budget Office, 2001:11-12).²⁶ Electricity wholesale prices began to rise above historic levels in May 2000. The first serious incident that indicated that California’s electricity market had run into trouble occurred on May 22. The ISO had anticipated an electricity shortage in its reserves, and declared a Stage 2 emergency at 11:40 a.m. The Stage 2 emergency indicated that the ISO’s electricity reserves had fallen below 5%. During the previous months, these notifications and the subsequent measures that were attached to the notifications had proven effective in mitigating energy reserve problems and helped to restore prices to normal levels. But this time, electricity wholesale prices did not drop and stayed at extraordinary high levels for the rest of the summer.

On June 14, 2000, PG&E—for the first time in its history—had to interrupt its services, in this case to almost 100,000 customers in the San Francisco Bay Area.²⁷ The precipitating problem, although unrelated to the prices in the electricity market, brought home the increasing stress and strain under which California’s high-voltage electricity grid was operating as a result of restructuring. The high temperatures of the day, combined with a disproportionate number of local generation units offline along with the inability to import enough electricity through to a lack of transmission capacity in the Bay Area, resulted in three hours of rolling blackouts. Throughout the 2000 summer, wholesale electricity prices in California were nearly 500% higher than during the same months in 1998 and 1999 (Joskow, 2001a:1). By July 2000, the ISO’s independent Market Surveillance Committee, charged with the review of California’s electricity market performance and rules had concluded:

California’s energy and ancillary services markets have not been workably competitive during the last two summers...[We] are unable to conclude that California’s energy and ancillary

²⁶ Hydropower is often used to provide “peak” power that is used for both grid management and for generating electricity during peak demand hours. The technological explanation for this use of hydropower is the comparatively rapid reaction time of these power plants to voltage changes. This flexibility contrasts with the slower reaction of both nuclear and fossil-fuel plants run on steam. The latter types, as they need relatively more time to react to changes, are generally used to provide so-called base-load power: a constant output of electricity according to pre-arranged schedules. Because grid management and energy demand vary enormously during summertime peaks, large amounts of hydropower are used during the summer to meet these contingencies. The result was that “15 to 20% of the electricity California needed for the summer was not available from traditional out-of-state sources” (California State Auditor, 2001:59).

²⁷ See Kahn & Lynch (2000:22) for a detailed specification of the number of customers and amount of load that was shed.

services markets will be workably competitive during high-demand periods this summer (Kahn & Lynch, 2000:45).

The first to be really directly affected by hyper-prices in the wholesale market, were, however, SDGE's customers. As noted a moment ago, SDGE was allowed to pass on its electricity wholesale prices to customers, causing electricity bills to nearly double compared to a year before. In September 2000, state legislators responded to public outcries and reinstalled a retail price cap, albeit SDGE incurred considerable losses when it had to buy electricity at hyper-prices (Congressional Budget Office, 2001:21). During the summer of 2000, the other two IOUs reported huge losses as well because they were obliged to buy power at wholesale prices far higher than the retail rates against which electricity could be sold.²⁸ The CPUC ignored the utilities' requests for retail rate increases during the summer months.

Even during times of reduced electricity demand and lower temperatures, electricity wholesale prices remained above average. Between May and December 2000, the average electricity prices traded in the PX were "between 2 and 13 times higher than in the same months of the previous year" (General Accounting Office, 2001:3). At the same time natural gas prices rose throughout the country, but especially in California. Both electricity wholesale and gas prices remained high throughout the year 2000. Consequently, PX trading suffered. The ISO found itself increasingly having to purchase electricity in the ancillary services market to serve California's load as both electricity generators and electricity purchasers (the utilities) engaged in strategic behavior in the electricity markets (California State Auditor, 2001). On August 1, 2000 more than 20% of California's load had to be met by real-time purchases of electricity by the ISO to keep the lights on (Faruqui *et al*, 2001:41). In effect, the ISO/PX markets had moved far from the intended market design by the summer of 2000, as can be seen in Figure A-4.

In Figure A-5, it is evident that the monthly average wholesale electricity price had risen to over \$250 per MWh by December 2000. Natural gas prices in California reached their maximum on December 8-10, 2000, \$58.76/MMBtu in Southern California.²⁹ At such high prices for natural gas, many generators claimed they could not profitably generate energy and sell it at or below the established price cap (Detmers, 2001). As a result, the financial condition of the utilities deteriorated rapidly. Late 2000, California's wholesale electricity markets had broken down with most of California's power being purchased outside these markets in bilateral deals at prices higher than previously declared price caps. Below Figure A-4 and A-5, Table A-3 shows the explosion in out of market purchases.

²⁸ By the autumn of 2000 PG&E and SCE had debts amounting to \$5 billion (Faruqui *et al*, 2001:37).

²⁹ Compared to a national high of \$10.52 for December 2000 (Detmers, 2001:2).

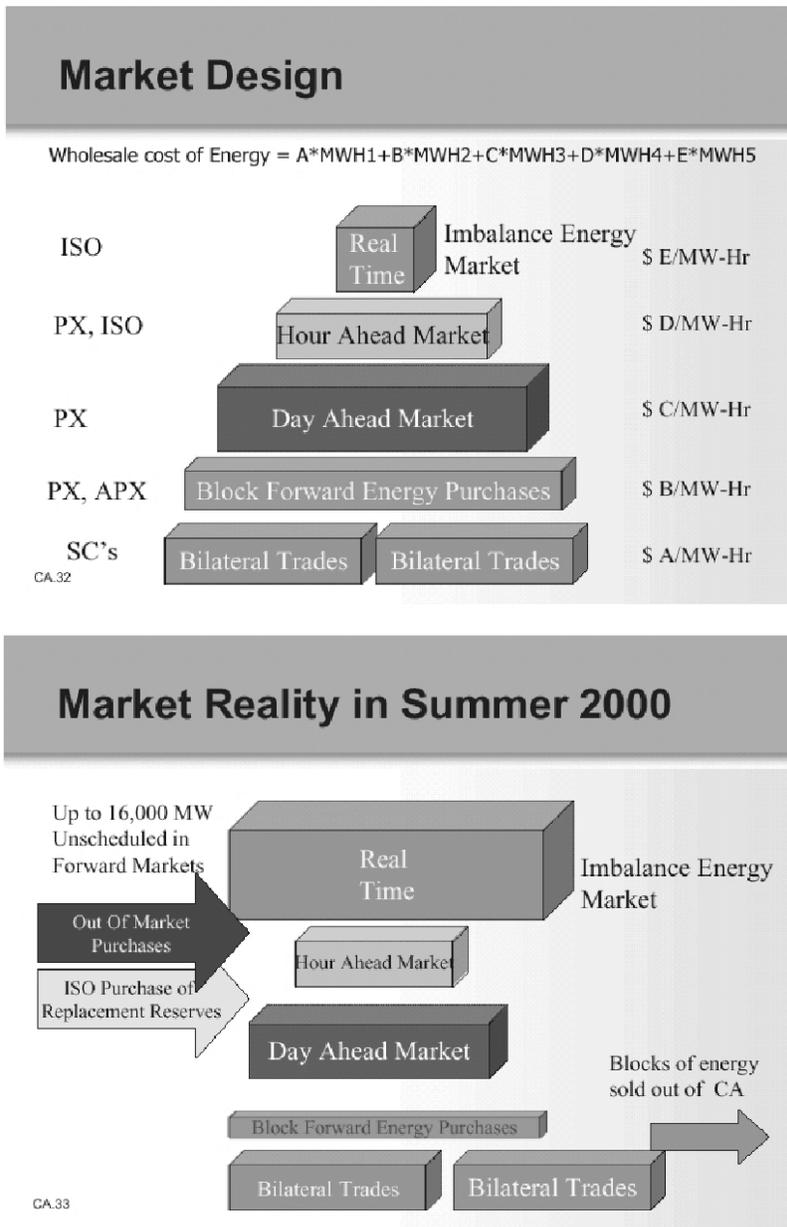
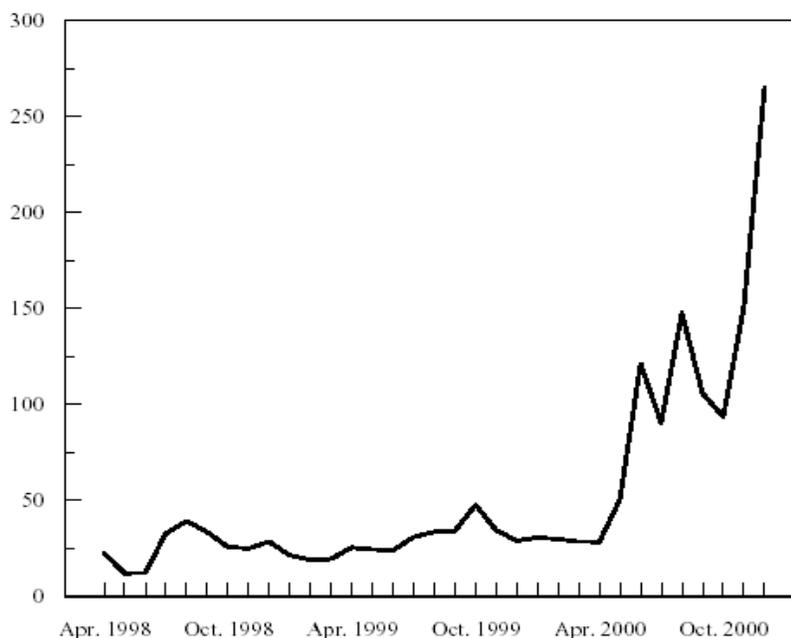


Figure A-4
California's Market Design and Market Reality in 2000
 Source: Lee (2001), pp. 4 and 6, Figures 5 and 9



April 1998 through December 2000

Figure A-5
Average Prices (in Dollars) Utilities Paid for Electricity in the Cal PX Day-Ahead Auctions
 Source: Congressional Budget Office (2001), pp.18, Figure 4

Table A-3
Volume Purchased Out-of-Market June through December 2000 (in Megawatts)

Month	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Average hourly out-of-market MWhs purchased (hours 12-19 only)	26,880	79,205	46,872	45,150	40,796	208,950	487,382

Source: California State Auditor (2001), pp. 29, Table 2

By December 2000, the UDC's unpaid accounts reached billions of dollars and the creditworthiness of PG&E and SCE were now in question. That month, the utilities stopped payments to the ISO and some small generators. The smaller generators ran up against their credit limits and stopped selling electricity to California. Even though the CPUC approved a 10% electricity retail price increase by early January 2001, the 1 cent/kWh surcharge was far too little for the utilities to cover their ongoing wholesale power costs, let alone to cover their previously acquired debts. The ISO was as well hobbled by the inability of the utilities to pay. ISO income was largely dependent on the utilities' payments for the use of the high-voltage electricity network and for ISO congestion management, such that now the ISO also became financially non-creditworthy. The potential repercussions for service and grid reliability were extreme, as the ISO was the supplier of last resort in order to ensure load and generation were

balanced in the California electricity system. The ISO faced buying the generation needed to maintain system reliability while claiming the costs from the now nearly bankrupt utilities.

Increasingly, electricity producers refused to sell electricity to both the utilities *and* the ISO, preferring to sell their electricity in other electricity markets and other states. An accumulation of cold weather and short hydroelectric power supply in a simultaneous strong need for electricity in the Pacific Northwest (Detmers, 2001:3). A number of federal emergency orders by the Secretary of Energy were barely sufficient to keep California's lights on in December and January. Through December and January, the Secretary of the Treasury tried to facilitate a negotiated settlement with the Governor of California, California utilities, and the major independent generators and marketers. His efforts were unsuccessful (Joskow, 2001b:43). Prices rose even further and rolling blackouts became inevitable.

On January 11, 2001, California was hit by a winter storm and the ISO had to declare a Stage 3 emergency, as its power reserves had fallen below 1.5%. Rolling blackouts were narrowly avoided that day through some last minute emergency electricity purchases. By mid-January the utilities had run out of cash and stopped paying their bills for power they had already purchased. The credit status of PG&E and SCE continued to spiral downward and on January 16 and 17, 2001, respectively, both Moody's and Standard & Poor's lowered the credit ratings of PG&E and SCE to junk status. This triggered even greater reliability and market stability concerns. FERC directed the ISO to ensure the presence of a creditworthy counter party to ensure financial backing for all third party energy procured for PG&E and SCE through the ISO markets (Detmers, 2001:i). Most of California's electricity trading was conducted outside of the PX, because no one was interested in dealing with the almost bankrupt PG&E and SCE.

Under increasing public pressure, California's legislature, the CPUC and Governor Gray Davis decided to involve the state in the electricity crisis.³⁰ As a stop-gap measure to relieve California's electricity shortages, Davis ordered the state to purchase the electricity the utilities needed in the wholesale market and in turn sell it to the indebted utilities against regular costs. On January 17, 2001, the State of California, acting through the California Department of Water Resources (CDWR), began purchasing energy on behalf of the UDCs, using, for this purpose, an appropriation of \$400 million from the State's General Fund.³¹ However, this did not prevent California from being hit by rolling blackouts due to a shortage of electricity, affecting some 380,000 customers on January 17 and 18.

On January 19, 2001, Governor Davis established the Department of Water Resources Electric Power Fund (DWREPF), temporarily authorizing CDWR to buy and sell electricity. On February 1, 2001, the Governor signed a bill, authorizing the department to enter into long-term contracts for the purchase of net short electric power prior to January 1, 2003. The bill continued the DWREPF on a permanent basis and mandated the DWREPF, in conjunction with State Treasurer, to issue revenue bonds for long-term financing of the DWREPF. Within CDWR, the California Energy Resources Scheduling Division (CERS) was set up to undertake these actions.

³⁰ See California State Auditor (2001), Appendix A for an overview.

³¹ "As a condition of its role as guarantor, CDWR had demanded non-public information from the ISO and special treatment not authorized in the existing ISO Tariff" (Detmers, 2001:i).

Throughout the first quarter of 2001, electricity supplies remained very tight, causing the ISO to issue a constant emergency status of grid reserves and the reliability of the provision of electricity. During the first two months of 2001, the ISO remained in a Stage 3 emergency for most of the time, so as to be able to procure and receive electricity from generators and neighboring electricity grids. As can be seen in Table A-4, the total number of declared emergencies—Stage 1, 2 and 3—rose dramatically. The ISO engaged in exchanging off-peak-hour energy for peak-hour energy with the Bonneville Power Administration (BPA) (Detmers, 2001:3). In total, the ISO had to drop firm load and institute rolling blackouts on 6 days, totaling 30 hours, with shortfalls ranging from 300 to 1,000 MW (Table A-5).

Table A-4
Comparison of California's Stage 1, 2 and 3 Emergencies

	Stage 1 (< 7% OR)	Stage 2 (< 5% OR)	Stage 3 (< 1.5% OR)
1998	7	5	0
1999	4	1	0
2000	55	36	1
2001	70	65	38

OR = Operating Reserves
Source: Detmers (2001), pp. 9

On 31 January 2001, the electricity crisis produced the first institutional victim in the restructured electricity sector: The PX suspended trading operations (California State Auditor, 2001:40; Congressional Budget Office, 2001:8). This, along with other developments, required the ISO to:

... make a series of emergency decisions in real time to maintain the grid while [and at the same time] uncertain whether dispatch instructions would be followed, that bids accepted by the ISO would be paid for and whether the ISO Tariff could encompass the demands on the ISO by FERC, federal and state courts, market participants and CDWR (Detmers, 2001:ii).

Table A-5
California's Rolling Blackouts

Date	Amount of electricity	Total MWs
January 17	500 MW for three hours	1,500 MW
January 18	1,000 MW for 1 hour, 500 MW for 2 hours	2,000 MW
March 19	500; 1,000; 800; 800; 401; and 401 MW	3,902 MW
March 20	500; 500; 300; 300; 300; and 300 MW	2,200 MW
May 7	300; 300; 300; 600; 600 and 300 MW	2,400 MW
May 8	300; 300; 300; 300; 400 and 400 MW	2,000 MW
Total		14,002 MW

Source: Hawkins, D. (2001), 'The California Report', presentation, October 2, 2001

Because of its massive power purchases to cover the loads California's almost bankrupt utilities, CERS, through CDWR, soon established itself as one of the largest electricity buyers in the nation, purchasing about 90% of the utilities' wholesale electricity needs, which amounted to about one third of California's total power use (Congressional Budget Office, 2001:28). Just from January to May 2001, the state of California paid an estimated \$8 billion to keep California's lights on, severely tapping into its financial reserves (Joskow, 2001a). To reduce the volatility of the electricity prices and the amount of electricity, the state had to buy electricity initially out-of-market and later on in the day-ahead and real-time electricity markets. CERS further began to negotiate large forward electricity contracts against set rates, committing the state to long-term electricity contracts. The contracts were at the time thought to amount to between \$40 billion and \$50 billion in the coming decades. At the time of writing, some of these contracts have been and are being renegotiated. To cover the costs of providing electricity to California, the state has plans to issue \$11.5 billion in bonds.

By the spring of 2001, the CPUC had raised the average retail rates some 40% to help the utilities cover their large electricity wholesale costs. Again, it was too little too late. In April 2001, PG&E, the largest of the three investor-owned utilities, filed for bankruptcy, claiming debts of \$8.9 billion (Congressional Budget Office, 2001:17).

As the summer of 2001 approached, it was feared that California would again be in serious trouble. Analysts expected peak demand to exceed supply by 5,000 MW and the ISO expected California would have more than 30 days of rolling blackouts, affecting as many as 5 million people (Faruqui *et al.*, 2001:47). FERC predicted the state could suffer as much as 260 hours of rolling blackouts.

To prevent a second series of rolling blackouts in the summer of 2001, FERC issued new orders curtailing the freedom of both generators and utilities in the market. During the summer,

electricity supply shortages did not occur on the scale initially predicted. California's peak-load reduction, market redesign and repeated demands for energy conservation seemed to pay off as electricity demand reduction and conservation peaked during the summer months as can be seen in Tables A-6 and A-7. From 1998 to 2000 California's total energy consumption during the summer months had steadily increased (81.4 TWh, 82.3 TWh and 87.3 TWh respectively), but this dropped significantly in the summer of 2001 (82 TWh).³² The ISO further pointed at the softening economy and mild weather as responsible for the decline.

Table A-6
California's Load Growth and Operating Reserves (in Megawatts)

	1996	1997	1998	1999	2000	2001
<i>Peak load (in MWs)</i>	47,331	49,493	50,049	51,287	49,097	48,597
<i>15 % Operating Reserve (in MWs)</i>	3,786	3,959	4,004	4,103	3,928	3,888
<i>7% Operating Reserve (in MWs)</i>	3,313	3,465	3,503	3,590	3,437	3,402

Source: Elliott, T. (2001), 'Summer 2001 "What Happened"?, Winter 2001/02 Outlook', presentation, October 8

Table A-7
Conservation in California During the Summer of 2001 (in Megawatts)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
<i>Peak Day Forecast</i>	33,743	32,195	32,233	31,888	34,657	39,637	41,599	42,528
<i>Peak Day Actual</i>	31,652	29,617	29,266	29,022	31,062	34,067	37,144	38,732
<i>Delta</i>	2,091	2,578	2,967	2,866	3,595	5,570	4,455	3,796
<i>Reduction</i>	- 6.2%	- 8.0%	- 9.2%	- 9.0%	-10.4%	-14.1%	-10.7%	- 8.9%

Source: Elliott, T. (2001), 'Summer 2001 "What Happened"?, Winter 2001/02 Outlook', presentation, October 8

During the summer months, electricity wholesale prices dropped substantially, falling below the prices CERS had negotiated in its long-term electricity contracts. In July 2001, average costs of the state's power purchases (\$133 MWh) were far above the real-time spot price of electricity (\$82 MWh) (Congressional Budget Office, 2001:30).

³² This was 5.3 TWh less than the total consumption during the summer of 2000. One TWh is equivalent to 1 x 10¹² watt-hours. The energy reduction seen during the summer of 2001 would be sufficient to serve approximately 350,000 residential homes for an entire year (Detmers, 2001:5).

The result of California's restructuring process has been to raise electricity rates to unprecedented levels and have left the restructured electricity market in a shambles. At the time of writing, the effects of the electricity crisis are from over. However, the most visible sign of the electricity crisis—the rolling blackouts—have been avoided for reasons outlined in the main text of this report. In total, California suffered 'only' 30 hours of rolling blackouts, totaling some 14,000MWh of load. When compared to the electricity usage of California households, one MWh will normally suffice to provide between 600 to 1200 homes with power (Brown & Koomey, 2002:4). Assuming one MWh would be enough to provide 750 homes with power, the total amount of load shed during the crisis equals a blackout of less than one hour for all of California's 10.5 million households.³³

With the bankruptcy of PG&E and the PX, and the emergence of CERS and the state's commitment to a long-term involvement in electricity contracts, California's electricity industry once again has been "restructured." CDWR replaced the utilities as the buyers of wholesale power in the markets and limited their role to the distribution of electricity. As a result, competition in the electricity market was severely hampered. Not only has CERS gained a virtual monopoly over electricity demand, the state also gained the authority from the legislature to seize generating power. CPUC's regulatory role was subsequently restricted as CERS falls outside the CPUC's authority.

The most recent institution to be added to California's electricity industry—but doubtless not the last for reasons outlined in the report—has been the California Consumer Power and Conservation Financing Authority or California Power Authority (CPA). Created during the summer of 2001 as the result of "intense executive and legislative activity to restore stability to the energy market, the CPA was envisioned to serve as a vehicle in acquiring additional power to meet the energy needs of California's consumers and to provide insurance for the State by securing a sufficient reserve of power." (CPA, 2002:1). To do this, the Power Authority is committed to the financing of renewable energy sources and peaker plants (about 3,000 MW) to reduce California's dependence on the spot power markets and restore California power reserve capacity to 15%, the amount previously prescribed by the IOUs. As such, the Power Authority enters as a new and important actor in a continuously "restructured" electricity sector envisioned by no one less than a decade ago.

Let us now turn to a concluding discussion of the operations in and around the ISO control room, the current focal point of the high reliability network in California.

Operations In and Around the Control Room of the California Independent System Operator (ISO)

The ISO employs hundreds of staff, but arguably its most important are those in and around its control room for ensuring service and grid reliability. In the middle of the ISO complex, its location symbolizes its central importance.

³³ The number of households was taken from Census 2000 Summary File 1[California]/prepared by the U.S. Census Bureau, 2001 at the time of writing to be found at: www.countingcalifornia.cdlib.org.

The ISO control room monitors and manages the more 239 billion kilowatt hours of electricity that is annually provided by over 1,100 generation units and sent through more than 25,500 miles of high-voltage transmission network, covering some 124,000 square miles – about 75% of California’s territory (ISO:2001). It is an always-on management, 24 hour, 7 days a week, 365 days a year. At the time of writing, the ISO employed 42 control room operators, 25 are located in the Folsom control room and 17 in its Alhambra facility in southern California. The Alhambra control room normally augments Folsom’s operators and is directly responsible for certain activities (e.g., the pre-scheduled Reliability-Must-Run (RMR) units, controlling and monitoring California’s southern transmission grid and interconnections with adjacent control areas). If necessary, Alhambra can completely take over Folsom’s control room operations.

Access to the 15,000 square foot control room at Folsom is highly restricted. Crews in the ISO control room (hereafter we will be talking exclusively of Folsom operations, unless otherwise indicated) typically consist of 15 control room operators who work in 12 hour shifts.³⁴ Table A-8 provides an overview of the functions within the control room teams.

Table A-8
An Overview of the Functions Within the ISO Control Room Teams

Folsom Control Room	Alhambra Control Room
Shift Manager	Generation Dispatcher
Generation dispatchers (2)	Real-Time Schedulers (2)
Real-Time Grid Resources Coordinator	Transmission Dispatchers (2)
Hour-Ahead Grid Resources Coordinator	
Real-Time Schedulers (2)	
Day-Ahead Grid Resources Coordinator	
Transmission Dispatchers (2)	

Source: Price Waterhouse Coopers (2001), *CAISO Operational Audit, November 20 2001*, <http://www.caiso.com/docs/09003a6080/13/e2/09003a608013e20f.pdf> (June 2002), Diagram 1, pp. 66

The control room is semi-circular in shape, the most notable feature being a wall-to-wall mosaic map-board giving a schematic overview of California’s high voltage electricity grid. Small indicators show the status of important circuits, power plants and substations.³⁵ Centrally located above the mosaic board a ticker tap monitor displays various prices and load numbers. The left-hand side of this room (when facing the front map-board) is occupied by a large, multi-functional video-monitor, which displays the status of the most important indicators and parameters of California’s electricity grid every four seconds provided by the Energy Management System

³⁴ The staff numbers vary during day and night operations. The day-ahead market for example does not operate at night.

³⁵ <http://www.caiso.com/aboutus/infokit/ControlCenter.html>

(EMS). Above the video-monitor a small panel shows the frequency in California's electricity grid, one of the most important control variables the ISO has to operate the grid.

A final characteristic element of the ISO's Folsom control room is its meeting room. This room contains a large glass wall that provides an excellent view of what is going on in the control room and video-monitor. This room is called the "fishbowl" for obvious purposes. The fishbowl is equipped with all sorts of equipment that allow it to be used as a conference room for emergency conference calls and meetings of the support staff during emergencies.

The room has 13 computer consoles for the control room operators. Each console is made up of a series of monitors and computers to provide the control room operator with the software to perform his or her function. During hurried periods of the day, operators are often on the phone, when not looking at their monitors. An overview of the ISO control room is provided by Figure A-6.

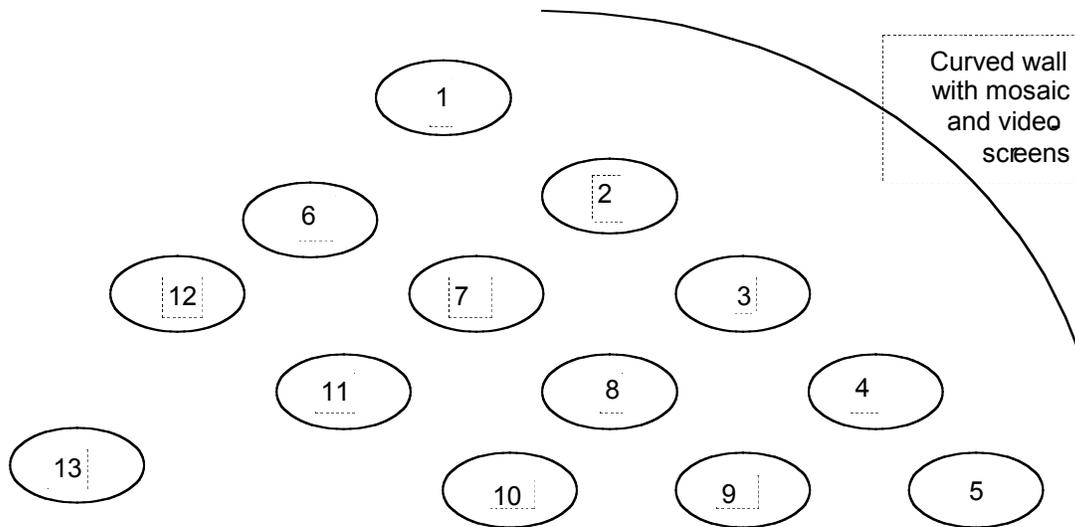


Figure A-6
A Schematic Display of the ISO's Folsom Control Room

What we have numbered as console 1 is occupied by the WSCC California power area security coordinator, who is responsible for monitoring this part of the western grid.³⁶ This western grid includes not only California, but comprises the western US and parts of Canada's and Mexico's high-voltage electricity grid. This part of the US grid is stability limited, which means that the system has inherent properties that make it unstable after certain limits (maximum load on transmission lines and substations) are violated. To prevent local or regional outages from cascading and bringing down the entire western grid, the system is equipped with Remedial Action Schemes (RAS) that will automatically trip and island the grid into smaller subsystems once critical loads have detected on important parts of the system. The maximum performance criteria for all important elements of the transmission grid are further documented and imposed as operating limits to prevent this "islanding".

³⁶ Officially, the WSCC security coordinator is not considered part of the ISO's control room team. These functions are funded by the WSCC.

To further ensure the reliable operation of the western electricity grid, the WSCC has issued the so-called Minimum Operating Reliability Criteria (MORC). The WSCC MORC's apply to all system operators under all conditions. The security coordinator ensures that the system operators in his/her region meet the various NERC and WSCC reliability standards and operates the grid within the allowed performance envelopes. Violations must be recorded and are subject to monetary penalties by the WSCC. The security coordinator also acts as liaison and coordinator between the ISO and neighboring system operators to respond to large-scale problems that (potentially) threaten the stability of the grids of multiple system operators.

The transmission dispatchers, who monitor and manage the status of the ISO's high-voltage transmission grid, occupy consoles 2 and 3. Changes in transmission directly affect the load constraints and the performance characteristics of the rest of grid. Transmission dispatchers process and coordinate responses to any threats to the transmission system, such as "overloading" of transmission lines along with planned or unplanned transmission outages. All changes in the status of (major components of) the grid such as outages of power plants or failures of transmissions line are processed and logged in the Scheduling Logging for the ISO of California (SLIC)-system by ISO staff.

At the time of our observations in the ISO, console 4 was occupied by CERS personnel, who purchased electricity on behalf of the state for distributed by the utilities.³⁷ Console 5 was used as an emergency support, back-up and training desk to ensure the ISO could accommodate future changes.

The ISO Control Room Meets the Electricity Market

Consoles 6 and 7 represent the day-ahead and hour-ahead market desks respectively. The ISO performs the role of a clearinghouse in three competitive electricity markets that exist apart from California's bilateral wholesale electricity market, i.e., the aforementioned congestion, ancillary services, and real-time imbalance markets. To repeat, these markets allow the ISO to "control and operate" the electricity system by continuously balancing load and generation and by providing the necessary resources to meet any potential threat to service and grid reliability. The market operations are therefore closely interwoven with the ISO's system control operations. Indeed, in the original design of restructuring, market transactions were to guide and coordinate grid management. The congestion market and the ancillary services market are both run in a day-ahead market, and a sequential hour-ahead market. The real-time imbalance market (Console 10, more in a moment) is run in real time.

The congestion market tries to ensure the efficient and optimal use of the high-voltage electricity grid, by comparing SC's filed balanced generation or demand schedules with the physical capabilities of the ISO transmission grid. The day-ahead grid resource coordinator collects these schedules and uses a congestion management program to determine if the schedules can be met given the physical constraints of the high-voltage electricity grid. If the grid cannot allocate the schedules, transmission lines will be congested. The ISO tries to facilitate both control of the

³⁷ In December 2001, FERC ordered the ISO to remove CERS out of the control room and are now considered as a scheduling coordinator (SC).

network and the efficient allocation of electricity by accepting “adjustment bids” to sell or buy more power to reduce congestion or charges if market participants continue to schedule power over transmission lines that are heavily used. A considerable amount of gaming has occurred in the congestion market, as witnessed by the Enron memos recently released by FERC.

The ancillary services market provides the ISO with the necessary resources to meet reliability standards for the grid. Generators with power plants with specific attributes (mainly reaction speed) are paid for bidding in this power and are further paid when their energy is bought by the ISO. NERC and WSCC reliability criteria determine the minimal amount of ancillary services the ISO is required to purchase. One of the WSCC/MORC-standards for example sets the system operator’s Operating Reserves at a minimum of 7% of it’s total load.

The day-ahead resource coordinator (at Console 6) facilitates the trading of electricity in the day-ahead market for both congestion management and ancillary services for the next day. The electricity schedules are aggregated, processed, compared with the electricity grid constraints, adjusted and redistributed to the market participants as the Final Day Ahead Schedule. The day-ahead market is meant to close at 10 a.m. each morning for the day ahead, though it has been left open later in order to allow late schedules.³⁸

Next to the day-ahead desk is the grid resource coordinator for the hour-ahead market (Console 7). This market allows market parties and the ISO to adjust the amount of electricity they have purchased in both the ancillary services and congestion markets in light of any changes since the day-ahead market was closed. In reality, the hour-ahead market is a misnomer, as it procures electricity three hours before actually dispatching the energy. The extra time is needed to allow the ISO’s computers to perform the highly complex task of (re)computing, (re)calculating and adjusting the electricity schedules and congestion management charges.

Consoles 8 and 9 are staffed by ISO’s real-time schedulers, but unofficially referred to as scheduling coordinators³⁹. ISO scheduling coordinators are responsible for monitoring, coordinating and managing the real-time electricity flows on California’s high-voltage transmission interconnections with neighboring electricity systems (interties). The loads and changes on these ties (and thus loads and flows in other grid areas) directly affect the stability of California’s electricity grid and need to be matched with California’s internal electricity flow patterns to avoid congestion and system disturbances. There are 25 such interties connecting the California grid with adjacent grids, the most important being California-Oregon Intertie (COI) that connects California to large hydropower resources in the northwest.

Apart from the constraints that the interties and other control areas pose, grid operations in California are further influenced by the inability to transport large volumes of electricity loads between different parts within California. A lack of transmission capacity inside California severely curtails the ability of the ISO to maintain the reliability of California’s electricity grid. By far the most important bottleneck in California’s electricity network today is the set of

³⁸ Until the issuing of the Final Schedule at 1 p.m., SC’s are offered a chance to readjust congested schedules to better fit the transmission constraints of California’s electricity grid.

³⁹ These control room staff are not to be mistaken for *commercial* scheduling coordinators (SC’s), intermediaries who file electricity schedules of aggregated electricity contracts to the ISO.

transmission lines designated as “Path 15”. Path 15 represents the only way to transport large volumes of electricity between north and south California and is wholly inadequate to perform that task. Problems are compounded by the relative uneven distribution of both the type and number of power plants and the distribution of load between the north and the south, which make the path critically vulnerable. During periods of congestion and severe strains on the grid, Path 15 is not able to transfer the amounts of electricity needed between these two areas. When this situation occurs, California’s electricity grid has to be managed as it were two independent grids, called North of Path 15 (NP15) and South of Path 15 (SP15). This means that electricity prices between the different zones can differ substantially. California has more such paths that are curtailing the options of control room operators to manage the flow of electricity over California’s electricity grid.

Even after the adjustments made in light of scheduled imports over the interties and in the day-ahead and hour-ahead markets, control room operators are compelled to make last minute changes and purchases so as to balance generation and load in real time. Electricity demand may be higher or lower than anticipated before the hour, transmission lines may trip, and power plants may suddenly go off-line. The differences are accounted for in the real-time-imbalance market, also known as the spot-market (Console 10). The real-time market is run by the real-time grid resources coordinator, commonly referred to within the ISO as the BEEPer. Each hour is divided in six ten-minute Balancing Energy and Ex-Post Price (BEEP) intervals in which a price is set to encourage market parties to either increase (incremental price, inc) or decrease (decremental price, dec) their power output. Bids are submitted for incing and decing, and it is the responsibility of this desk’s real-time grid resources coordinator, the BEEPer, to maintain that bid stack and process the inc or dec bids on the request of the generation dispatcher. The ISO generation dispatcher, commonly known as the gen dispatcher (Console 11), manages the grid in real time by estimating how much incing or decing is needed to control the Area Control Error (ACE), which shows the relative balance between generation and load in California’s grid. Maximum fluctuations in the ACE are set by WSCC reliability criteria. It is the task of the generation dispatchers to keep the imbalances of the ISO’s grid within these bandwidths. How well the generation dispatcher does his or her work, given the constraints of load and generation they operate under and the resources they have available in the BEEP stack (the list of bids in the BEEP market prioritized according to price) and through the interties, determines the number of violations the ISO faces.

Most control area reliability and performance standards are set by the WSCC and NERC.⁴⁰ The most important monitoring standards to assess the reliability of the ISO system operations are known as two control performance standards (CPS’s), set by NERC. Both standards require the ISO to control the frequency of the grid, the speed with which electricity flow alternates its current within one minute. This frequency standard within the U.S. has been set to 60.0000 Hertz and is an inherent characteristic of the stability of the grid. All efforts of system operators are aimed at controlling this frequency within stringently defined limits through the balancing of load, generation and interchange schedules.

⁴⁰ See for example NERC and WSCC standards. At the time of writing these could be found at: <http://www.caiso.com/docs/09003a6080/13/e9/09003a608013e91e.pdf>.

The first control performance standard (CPS1) compares the relationship of the frequency of the grid to the ACE. For example, if the frequency is above 60 Hertz and the ACE is positive (i.e. the control area has more generation than load and is “overgenerating”), then the system operator is contributing to the frequency problem and further driving up the frequency of the system. On the other hand, if the frequency is above 60.0000 Hertz but the ACE is negative (i.e. the control area has more load than electricity) it is trying to push the frequency problem back and contributing to the stability of the system. This makes CPS1 a reliability tool for the behavior of system operators in the interest of keeping the stability of the grid. However, the strict control standard (the monthly CPS 1 standard is set on 100%) is difficult to assess and monitor in real time. Violations usually are not known until long after the events took place in the control room.

The second control room performance standard, CPS2, is more directly related to the ISO’s operations. It makes use of the ACE as the crucial variable to monitor and limit excessive unscheduled power flows that could result from large ACEs and cause instabilities in the grid. CPS2 demands that the gen dispatcher minimize the fluctuations of the ACE within a maximum bandwidth every 10 minutes. The standards require a monthly 90% performance rate within these bandwidths, and allow the ISO a daily quota of 14 10-minute CPS2 violations. If the ISO’s monthly averages fall below 90%, the regional reliability organization will impose a fine.

Besides, the CPS violations, the ISO’s grid operations are constrained by other so-called Disturbance Control Standards (DCS’s) such as path-violations. Path violations record the crossing of safety-limits (measured both in load and in time) that have been set for the different transmission lines. Because the paths represent the vital links between the different grids, path violations are considered major violations and are penalized by the WSCC.

It is within these constraints that the generation dispatcher tries to control the grid. To keep the system within the above explained parameters, the generation dispatcher closely watches the frequency and ACE-trends, the output of all power plants and some vital path indicators on his monitor to determine the stability state and “movement” of the grid on the ACE and frequency parameters. To adjust frequency and ACE, the generation dispatcher can order contracted power plants to increase or decrease their electricity output. However, it takes time before the actual dispatching order of the generation dispatcher and the effects of his or her orders can be noted in the grid behavior. Power plants increase or decrease their power output rather slowly and it takes some time before the accumulative effects of these changes takes effect in the ACE and frequency rate.

Dispatching orders are sent through the Automatic Dispatch System (ADS) that notifies the power plant operators that the ISO has accepted their bids for real-time imbalance energy and asks them to execute their contracted commitment. The ADS system allows the power plant operator 120 seconds to react to accept or decline this offer, after which the next bid in the stack will be notified through ADS.

Besides the “manually operated” real-time market to cover for imbalances, the generation dispatcher also uses electricity from power plants that have contracted an amount of power plant output to the ISO for longer terms. The installation of a hard- and software package known as RIG’s (Remote Intelligent Gateways) allow the ISO’s Energy Management System (EMS) and

the generation dispatcher to directly control the output of these power plants. EMS helps the generation dispatcher by automatically dampening load differences, imbalances and other problems. The system that provides the ISO with automatic control over power plant output is known as Automatic Generator Control (AGC).

The gen dispatcher and BEEPer work closely together as a team. The inherent instability of electricity demand and behavior of the electricity system requires a constant fine-tuning and hands-on management of balance and load and therefore constant recourse to the real-time imbalance market. Next to the gen dispatcher is another gen dispatcher (Console 12), who assist the primary gen dispatcher and serves as a back up for communications, monitoring and assessment.

Finally, the grid operations shift manager completes the ISO's control room team. The shift manager is responsible for overall coordination within his or her team and acts as the liaison for that team with parties outside the control room. During each shift, each manager bears the ultimate responsibility for controlling the grid. All critical decisions, such as the issuing of emergency stages are taken by the shift manager in consultation with his or her superiors. As noted in the main text of this report, however, much of the activity in the control room is non-hierarchical and collegial, both during these emergencies and on more normal days.

All other departments within the ISO are critically related to the ISO control room operations. The following departments constitute the control room's wraparound as its direct support-network: Scheduling, Operations Support & Training, Market Operations, Operations Engineering, and Outage Coordination. These departments perform vital tasks that allow the ISO control room to function properly.

The scheduling department is responsible for the process called pre-scheduling, the pre-planning of generation and load schedules on the basis of the long-term and day-ahead markets. These schedules are also loaded into the computers for the hour-ahead market.

Operations support and training can provide the necessary personnel to back up to control room operators in times of emergencies and threatened system crashes. They are also responsible for the training of control room personnel.

Market operations designs and controls the software for the different electricity markets in which the ISO plays an active role. Changes in market rules, the ISO's Tariff, responses to strategic behavior and institutional changes as a result of the electricity crisis all had to be implemented in new software to ensure the proper working of the electricity markets and allow ISO control room personnel the ability to acquire the resources necessary to operate the grid. Updates in market software occurred on a regular, almost weekly basis during the crisis in 2000.

The operations engineering department undertakes in-depth studies of the vulnerabilities and operating characteristics of California's electricity grid. Using computers to model (parts) of California's grid under specific operating conditions (nomograms), operations engineers are able to anticipate and analyze the system's behavior to specific conditions. In emergency situations

and periods of heavy system loads, operations engineers will support the control room operators with their more in-depth and intimate knowledge of specific parts of the electricity grid.

Finally, outage coordination allows the ISO more influence and control over the planned and unplanned generation outages in California's restructured industry. Especially during the electricity crisis, but also in some constrained load areas, the ISO has at times lacked the resources to reliably operate the grid. Outage coordination helps distribute the maintenance of power plant operators throughout the year, which in turn, allows control room personnel to better cope with controlling the grid.

B

PRIMARY FEATURES OF HIGH RELIABILITY ORGANIZATIONS

High Technical Competence

HROs are characterized by the management of technologies that are increasingly complex and which require specialized knowledge and management skills in order to safely meet the organization's peak load production requirements (Rochlin 1993: 14; La Porte 1993: 1). What this means in practice is that the organizations are continuously training their personnel, with constant attention devoted to recruitment, training, and performance incentives for realizing the high technical competence required (Roberts 1988: Figure 3; La Porte 1996: 63). To do so means not only that there must be an extensive database in the organization on the technical processes and state of the system being managed, but that this "database" includes experience with differing operating scales and different phases of operation—the proposition being that the more experience with various operating scales and the more experience with starting and stopping risky phases of those operations, the greater the chance that the organization can act in a reliable fashion, other things being equal (La Porte 1993: 3; Perrow 1994: 218).

High Performance and Close Oversight

Technical competence in an HRO must be matched by continual high performance. The potential public consequences of operational error are so great that the organization's continued survival, let alone success, depends on reliably maintaining high performance levels through constant, often formal oversight by external bodies. As Rochlin (1993: 14) puts it, "The public consequences of technical error in operations have the potential for sufficient harm such that continued success (and possibly even continued organizational survival) depends on maintaining a high level of performance reliability and safety through intervention and management (i.e., it cannot be made to inhere in the technology)." Accordingly (Rochlin 1993: 14), "Public perception of these consequences imposes on the organizations a degree of formal or informal oversight that might well be characterized as intrusive, if not actually comprehensive." La Porte (1993: 7) adds, "Aggressive and knowledgeable formal and informal watchers [are] IMPORTANT. Without which the rest [i.e., high reliability] is difficult to achieve."

Note that "oversight" does not mean strict supervision of personnel within the HRO; in fact, overly close supervision is inimical to achieving high reliability (Schulman 1993a). Rather the oversight in question typically comes from external bodies that demand high reliability from the HRO's senior managers, who in response allocate resources to achieve that reliability.

Constant Search for Improvement

A feature related to high technical competence and constant monitoring is the continued drive to better HRO operations. Personnel constantly strive to improve their operations and reduce or otherwise avoid the hazards they face, even when—or precisely because—they are already performing at very high levels. "While [HROs] perform at very high levels, their personnel are never content, but search continually to improve their operations" (Rochlin 1993: 14). They seek improvement "continually via systematic gleaning of feedback" (La Porte 1996: 64). Notably, the quest is not just to do things better, but to reduce the intrinsic hazards arising from the activities of many HROs (La Porte 2000, personal communication).

Hazard-Driven Adaptation to Ensure Safety

HROs face hazards that drive them to seek flexibility as a way of ensuring safety. "The activity or service [of these HROs] contains inherent technological hazards in case of error or failure that are manifold, varied, highly consequential, and relatively time-urgent, requiring constant, flexible, technology-intrusive management to provide an acceptable level of safety [i.e., reliability] to operators, other personnel, and/or the public" (Rochlin 1993: 15). The more hazardous the operations, the more the pressure to ensure high reliability of those operations. Weick and Sutcliffe (2001: 14) characterize this flexibility as resilience: "The signature of HROs is not that it is error-free, but that errors don't disable it."

Highly Complex Activities

Not unexpectedly, the more complex the actual operations and activities performed, that is, the more inherently numerous, differentiated, and interdependent they are, the greater the pressure to operate in a highly reliable fashion (e.g., Rochlin 1993: 15). What this means in practice is that HROs often find it "impossible to separate physical-technical, social-organizational, and social-external aspects; the technology, the organization, and the social setting are woven together inseparably" (Rochlin 1993: 16). In such organizations, its technology, social setting, and units are extremely difficult to tease apart conceptually and practically. Such complexity characterizes many activities of many HROs. Note the qualification "many." Not all activities in an HRO are complex (La Porte 2000, personal communication). The point here is that the more complex (and the more hazardous) the operations of an organization in combination with the other features discussed above and below, the greater the pressure to manage in a highly reliable fashion.

High Pressures, Incentives and Shared Expectations for Reliability

HRO activities and operations must meet social and political demands for high performance, with safety requirements met in the process and clear penalties if not (Rochlin 1993: 15). One way to do so is to ensure that those who provide the services work and live close to the system they manage—they fly on the airplanes they build or guide, they live downwind of the chemical plants they run or on the floodplains they manage, and their homes depend on the electricity and water they generate (Perrow 1994: 218).

Culture of Reliability

Since the HRO must maintain high levels of operational reliability, and safely so, if it is to be permitted to continue to carry out its operations and service provision, a culture of reliability comes to characterize the organizations (Rochlin 1993: 16; Roberts 1988: Figure 3). This means in practice that the organizations often exhibit clear discipline dedicated to assuring failure-free, failure-avoiding performance (La Porte 1993: 7). Such a culture does not mean the organization is saturated with formal safety regulations and protocols, which as with overly close supervision end up working against the achievement of high reliability. A culture of high reliability is one in which core norms, values, and rewards are all directed to achieving peak load performance, safely, all the time, informally as well as formally (La Porte 1997, personal communication). As Rochlin (1993: 21) puts it, "the notion of safe and reliable operation and management must have become so deeply integrated into the culture of the organization that delivery of services and promulgation of safety are held equally as internal goals and objectives: neither can be separated out and 'marginalized' as subordinate to the other, either in normal operations or in emergencies."

Reliability Is Not Fungible

Because of the extremely high consequences of error or failure, HROs cannot easily make marginal tradeoffs between increasing their services and the reliability with which those services are provided (Rochlin 1993: 16). "Reliability demands are so intense, and failures so potentially unforgiving, that ...[m]anagers are hardly free to reduce investments and arrive at conclusions about the marginal impacts on reliability" (Schulman 1993b: 34–35). There is a point at which the organizations are simply unable to trade reliability for other desired attributes, including money. Money and the like are not interchangeable with reliability; they cannot substitute for it. High reliability is, in brief, not fungible.

Limitations on Trial-and-Error Learning (Operations Within Anticipatory Analysis)

Given the above, it is not surprising that HROs are very reluctant to allow their primary operations to proceed in a usual trial-and-error fashion for fear that the first error would be the last trial (Rochlin 1993: 16). They are characterized by "inability or unwillingness to test the boundaries of reliability (which means that trial-and-error learning modes become secondary and contingent, rather than primary)" (Rochlin 1993: 23). While HROs do have search and discovery processes, and elaborate ones, they will not undertake learning and experimentation that expose them to greater hazards than they already face. They learn by managing within limits and, if possible, by setting new limits, rather than testing those limits for errors (La Porte, personal communication).

As Rochlin (1993: 14) puts it, HROs "set goals beyond the boundaries of present performance, while seeking actively to avoid testing the boundaries of error" (Rochlin 1993: 14). Trial and error learning does occur, but this is done outside primary operations, through advanced

modeling, simulations, and in other anticipatory ways that avoid testing the boundary between system continuance and collapse.

Flexible Authority Patterns

HROs "structur[e] themselves to quickly move from completely centralized decision making and hierarchy during periods of relative calm to completely decentralized and flat decision structures during 'hot times'" (Mannarelli *et al.* 1996: 84). These organizations have a "flexible delegation of authority and structure under stress (particularly in crises and emergency situations)" (Rochlin 1996: 56; see also Roberts *et al.*, 1994), where "other, more collegial, patterns of authority relationships emerge as the tempo of operations increases" (La Porte 1996: 64). When this ability to rapidly decentralize authority under stress is combined with the persistent drive to maintain flexibility and high levels of competence in an HRO, the management emphasis is to work in teams based on trust and mutual respect (La Porte, 2000, personal communication). In this way, emergencies can be dealt with by the person on the spot, whose judgment is trusted by other members of the team that works together in these and less charged situations.

Positive Design-Based Redundancy

Last, but certainly not least, HROs are characterized by the multiple ways which they respond to a given emergency, including having back-up resources and fallback strategies. This can happen in three ways:

1. Functional processes are designed so that there are often parallel or overlapping activities that can provide backup in the case of overload or unit breakdown and operation recombination in the face of surprise.
2. Operators and first-line supervisors are trained for multiple jobs including systematic rotation to assure a wide range of skills and experience redundancy.
3. Jobs and work groups are designed in ways that limit the interdependence of incompatible functions (La Porte 1996: 63–64; see also Rochlin 1993: 23).

C

REFERENCES

Aldrich, Howard. *Organizations Evolving*. Sage Publications, Thousand Oaks, CA 1999.

Amin, M. "Complex Interactive Networks Workshop". EPRI Grid Operations and Planning. Washington, DC. 2001. Available online at:
http://www.epri.com/attachments/267936_CINS_wkshp.pdf

Batteau, A. (2001), "The Anthropology of Aviation and Flight Safety" in *Human Organization: Journal of the Society for Applied Anthropology*. Vol. 60, No. 3, pp. 201-211.

Brown, R.E. and J.G. Koomey (2002), "Electricity Use in California: Past Trends and Present Usage Patterns", Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, LBL-47992, forthcoming in: *Energy Policy*,
<http://enduse.lbl.gov/Projects/CAdata.html>, (June, 2002).

Bushnell, J. and S. Oren (1997), "Transmission pricing in California's proposed electricity market", in: *Utilities Policy*, Vol. 6, No. 3, pp. 237-244,
www.ieor.berkeley.edu/%7Eoren/pubs/buor97.pdf, (June, 2002)

California Power Authority (2001), *2001 Annual Book*, Sacramento CA,
www.capowerauthority.ca.gov/AnnualBook/2001AnnualBook.pdf, (June, 2002)

California Public Utilities Commission (1993), *California's Electric Services Industry: Perspectives on the Past, Strategies for the Future*, Division of Strategic Planning, San Francisco CA, www.ucei.berkeley.edu/ucei/Restructuring_Archive/Yellow_book.pdf, (June, 2002)

California State Auditor (2001), *Energy Deregulation: The Benefits of Competition Were Undermined by Structural Flaws in the Market, Unsuccessful Oversight, and Uncontrollable Competitive Forces*, 2000-134.1R, Sacramento CA, www.bsa.ca.gov/bsa/, (June 2002)

Congressional Budget Office (2001), *Causes and lessons of the California electricity crisis*, Congress of the United States, Washington DC, www.cbo.gov, (June, 2002)

Demchak, C. C. 1991. *Military organizations, complex machines: Modernization in the U.S. armed services*. Ithaca, N.Y.: Cornell University Press.

Demchak, C. C. 1996. Tailored precision armies in fully networked battlespace: High reliability organizational dilemmas in the 'information age.' *Journal of Contingencies and Crisis Management* 4 (2): 93-103.

References

- Detmers, J. (2001), *California ISO Operations Economic Book 2001*, Folsom CA, www.caiso.com, (June, 2002)
- Elliott, T. (2001), 'Summer 2001 "What Happened"?, Winter 2001/02 Outlook', presentation, October 8
- Energy Information Administration (1993), *The Changing Structure of the Electric Power Industry, 1970-1991*, U.S. Department of Energy, DOE/EIA-0562, Washington DC
- Energy Information Administration (1996), *The Changing Structure of the Electric Power Industry: An Update*, U.S. Department of Energy, DOE/EIA-0562(96), Washington DC
- Farrell, A. et al, "Bolstering the Security of the Electric Power System ", in *Issues in Science and Technology*, Spring 2002, pp. 49-56
- Faruqui, A. (et al)(2001), 'Analyzing California's Power Crisis', in: *The Energy Journal*, Vol. 22, No. 4, pp. 29-52
- Financial Times*, "Fake energy trades threaten free market.". May 16, 2002, p. 15
- General Accounting Office (2001), *Energy Markets, Results and Studies Assessing High Electricity Prices in California*, Book to Congressional Requesters, GAO-01-857, Washington DC, www.gao.gov, (June, 2002)
- Hawk, D.V. (1999), *Disconnect: A Transaction Cost Analysis of California Electric Power Industry Restructuring*, University of California, Berkeley
- Hawkins, D. (2001), 'The California Book', presentation, October 2, 2001
- Homer-Dixon, T. *Foreign Policy* January/February 2002, pp. 52-62
- ISO (2001), *California ISO Year 2000 Annual Book*, Conversations with the California ISO, Folsom CA, www.caiso.com, (June, 2002)
- Janis, I. 1982. *Groupthink*, 2d ed., Houghton Mifflin, Boston
- Joskow, P.L. (2000a), 'Deregulation and regulatory reform in the U.S. electric power sector', in: Peltzman, S. & C. Winston (Eds.), *Deregulation of network industries: What's next?*, AEI-Brookings Joint Center for Regulatory Studies, Washington, pp. 113-188
- Joskow, P.L. (2000b), 'Deregulation and regulatory reform in the U.S. electric power sector', revised discussion draft, February 17, econ-
www.mit.edu/faculty/pjoskow/files/BrookingsV2.pdf, (June, 2001)
- Joskow, P.L. (2001a), 'California's electricity crisis', NBER Working Paper Series, Working Paper 8442, www.nber.org/papers/w8442, (June, 2001)

- Joskow, P.L. (2001b), 'U.S. energy policy during the 1990s', paper presented for conference on American Economic Policy During the 1990s, John F. Kennedy School of Government, Harvard University, June 27-30, 2001, econ-www.mit.edu/faculty/pjoskow/files/usen1990.pdf, (December, 2001)
- Kahn, M., and L. Lynch (2000), *California's Electricity Options and Challenges*, Book to Governor Gray Davis, Sacramento, www.cpuc.ca.gov/published/book/GOV_BOOK.htm, (June, 2002)
- Kay, John. "A watchful eye on power: California, Auckland, Enron. They all prove the case for limited regulation of electricity," *Financial Times*, December 12, 2001 p. 14.
- Landau, M. 1969. Redundancy, rationality, and the problem of duplication and overlap. *Public Administration Review* 29 (July–August): 346–358.
- La Porte, T. R. 1993. "Organization and safety in large scale technical organizations': Lessons from high reliability organizations research and task force on 'Institutional trustworthiness,'" paper prepared for a seminar on Man-Technology-Organization in Nuclear Power Plants, Finnish Centre for Radiation and Nuclear Safety, Technical Research Centre of Finland, Olkiluoto, Finland. June 14–15.
- La Porte, T. R. 1996. High reliability organizations: Unlikely, demanding and at risk. *Journal of Contingencies and Crisis Management* 4 (2): 60–71.
- La Porte, T.R. 2000. Highly reliable Operations and the Rigors of Sustained Legitimacy: Matters of public trust and institutional constancy. Revision of paper presented to the workshop on "Organization Approaches to Organizational Reliability," Department Technologies et Sciences de l'Homme, Universite de Technologies de Compiègne, Compiègne, France, October 7-8, 1999.
- Lee, S.T. (2001), Analysis of the 2000-2001 California Power Crisis, Implications for Grid Operations and Planning, EPRI, March 12, Palo Alto CA, www.dvg-heidelberg.de/, (June, 2002)
- Lerner, A. W. 1986. There is more than one way to be redundant: A comparison of alternatives for the design and use of redundancy in organizations. *Administration & Society* 18 (3): 334–359.
- Lindblom, C. 1990 *Inquiry and Change: The Troubled Attempt to Understand and Shape Society*, New Haven, Conn., Yale University Press.
- Mann, C., "Special Report: Homeland Insecurity" in *The Atlantic Monthly* September 2002, Vol. 290, No. 2, pp. 81-102;
- Mannarelli, T., K. Roberts, and R. Bea. 1996. Learning how organizations mitigate risk. *Journal of Contingencies and Crisis Management* 4 (2): 83–92.

References

- March, J. G., L. S. Sproul, and M. Tamuz. 1991. Learning from samples of one or fewer. *Organization Science* 2 (1): 1–13.
- March, J. G. and J. P. Olsen. 1989. *Rediscovering institutions: The organizational basis of politics*. New York: Free Press.
- Marnay, C. (et al)(2001), “Direct Participation of Electrical Loads in the California Independent System Operator Markets During the Summer of 2000”, In: *Proceedings of the European Council for an Energy Efficient Economy Summer Study 2001*, Mandelieu, France, 11-15 June 2001, Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, LBNL-47897, eetd.lbl.gov/ea/EMS/books/47897.pdf, (June, 2002)
- Martin, Peter (2001), “Failure is always an option: Essential public services may be guaranteed by government—but their providers are not,” *Financial Times*, Thursday October 11, p. 14.
- Mensah-Bonsu, C. and S. Oren. 2001. *California Electricity Market Crisis: Causes, Remedies and Prevention*, CAISO.
- Meyer, Marshall W. and Lynne G. Zucker. *Permanently Failing Organizations*. Sage Publications, Newbury Park, CA 1989.
- Morgan, Gareth. *Images of Organization*, 2nd edition. Sage Publications Thousand Oaks, CA 1997.
- National Research Council. 2002. *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism*, Committee on Science and Technology for Countering Terrorism, Washington DC: National Academy Press.
- Perrow, C. 1984. *Normal accidents: Living with high-risk technologies*. New York: Basic Books.
- Perrow, C. 1994. The limits of safety: The enhancement of a theory of accidents. *Journal of Contingencies and Crisis Management* 2 (4): 212–220.
- Rochlin, G. 1993. Defining ‘high reliability’ organizations in practice: a taxonomic guide prologue. In *New Challenges to Understanding Organizations*, ed. K.H. Roberts, 11-32. New York: Maxwell Macmillan International.
- Roberts, K.H. 1988. *Some characteristics of high reliability organizations*. Berkeley, Calif.: School of Business Administration, University of California at Berkeley.
- Schiller, Ben (2002). “Long record of false starts: Electrical power companies around the world have considered the possibilities of powerline services, but relatively few of these ideas have become a reality,” *Financial Times*, March 20.
- Schulman, P.R. 1993a. The negotiated order of organizational reliability. *Administration & Society* 25 (3): 353–372.

- Schulman, P.R. 1993b. The analysis of high reliability organizations: A comparative framework. In *New challenges to understanding organizations*, ed. K. H Roberts. 33-54. New York: Maxwell Macmillan International.
- Senge, P. 1990. *The Fifth Discipline*. New York: Doubleday.
- Turner, B.A. 1978. *Man Made Disasters*. London: Wykenham Publications.
- Van Eeten, M.J.G. and E. Roe. 2002. *Ecology, Engineering and Management: Reconciling Ecosystem Rehabilitation and Service Reliability*. New York: Oxford University Press.
- Von Meier, A. 1999. Occupational cultures as a challenge to technological innovation. *IEEE Transactions on Engineering Management* 46:1: 104 –114.
- Weick, K.E. and K.M. Sutcliffe. 2001. *Managing the unexpected: Assuring high performance in an age of complexity*. San Francisco: Jossey-Bass
- Wildavsky, A. 1979. *Speaking Truth to Power*. Boston, Mass.: Little, Brown and Company.
- World Bank (2001), *The California Power Crisis: Lessons for Developing Countries*, Energy and Mining Sector Board, Washington DC, www.worldbank.org/html/fpd/energy/caexp.html, (June, 2002)