

DEVELOPING RESEARCH AND PRACTICE

“Going solid”: a model of system dynamics and consequences for patient safety

R Cook, J Rasmussen

Qual Saf Health Care 2005;14:130–134. doi: 10.1136/qshc.2003.009530

Rather than being a static property of hospitals and other healthcare facilities, safety is dynamic and often on short time scales. In the past most healthcare delivery systems were loosely coupled—that is, activities and conditions in one part of the system had only limited effect on those elsewhere. Loose coupling allowed the system to buffer many conditions such as short term surges in demand. Modern management techniques and information systems have allowed facilities to reduce inefficiencies in operation. One side effect is the loss of buffers that previously accommodated demand surges. As a result, situations occur in which activities in one area of the hospital become critically dependent on seemingly insignificant events in seemingly distant areas. This tight coupling condition is called “going solid”. Rasmussen’s dynamic model of risk and safety can be used to formulate a model of patient safety dynamics that includes “going solid” and its consequences. Because the model addresses the dynamic aspects of safety, it is particularly suited to understanding current conditions in modern healthcare delivery and the way these conditions may lead to accidents.

condition. The anesthetic was started because it had become routine to begin surgery in anticipation of resources *becoming* available rather than waiting for them to *be* available. The patient would have required an ICU bed for recovery after the procedure and the practitioners elected to halt the operation when it became apparent that no ICU bed would be available. Similar situations are reported by others (box 1). Such situations create opportunities for incidents and accidents.

We describe this shift in operations as *going solid*. “Going solid” is a nuclear power slang term used to describe a difficult to manage technical situation.³ Manageable behavior of a steam boiler depends on having both steam and liquid water present in the boiler. When a boiler becomes completely filled with liquid (goes solid), its operating characteristics shift suddenly and dramatically. The resulting situation is both hazardous and difficult to control. “Going solid” in hospitals creates problems that are, we believe, similar in many respects to those that accompany going solid in nuclear power plants. In particular, “going solid” in a hospital creates a series of critical relationships, tightly coupling the units of the hospital together so that events in one place have direct implications for the operations of all the others.

To describe, analyze, and assess this phenomenon, we introduce a model that explains how “going solid” works at a systemic level and its implications for the understanding of safety culture.⁴

DYNAMIC SAFETY MODEL

Rasmussen’s dynamic safety model (fig 1A) describes the feasible operating space for a sociotechnical system within three boundaries that form an envelope.⁵ This model is descriptive rather than normative. The operating point location is influenced by gradients that drive operations away from the workload and economic failure boundaries and towards the unacceptable performance (accident) boundary. Because the environment is dynamic, the operating point moves continuously; stability occurs when the movements of the operating point are small and, over time, random. Changes in the gradients (for example, increased economic pressure) move the operating point. The risk of an accident falls as distance from the unacceptable performance boundary increases. In practice, the precise location of the boundary of unacceptable performance is uncertain. Only accidents provide unambiguous information about its position.

Healthcare delivery systems routinely work at the very limits of their capacity. As with other expensive, complex domains, Hirschhorn’s “law of stretched systems” applies: “... we are talking about a law of systems development which is: every system always operates at its capacity. As soon as there is some improvement, some new technology, we stretch it ...”.^{1,2}

Practitioners are familiar with “bed crunch” situations where a busy unit such as a surgical intensive care unit (ICU) becomes the operational bottleneck within a hospital. Other units in the hospital usually buffer the consequences of a localized bed crunch by absorbing workload, deferring transfers, etc. We have observed situations where an entire hospital is saturated with work, creating a system wide bed crunch. The result is a dramatic change in workplace operational characteristics that creates new demands for coordination of work across the facility and increases the stakes for practitioner decisions and actions. To cite an example: a surgical procedure was cancelled after induction of anesthesia because a scheduled transfer of another patient out of the ICU was made impossible by deterioration of that patient’s

See end of article for authors’ affiliations

Correspondence to:
Dr R I Cook, 5841 S
Maryland Ave, MC4028,
Chicago, IL 60615, USA;
ri-cook@uchicago.edu

Accepted for publication
5 January 2005

Box 1 Example of a case of “going solid”

Setting

A large tertiary care facility in a major metropolitan area in the United States.

Event

Near the end of a routine scheduled surgical procedure on patient A, the circulating nurse called the recovery room in anticipation of bringing the patient to it. The recovery room placed the transfer from the operating room “on hold” because all the recovery room locations were filled by patients. Among these was patient B who should have been transferred from the operating room directly to an intensive care unit (ICU) bed. Patient B was in the recovery room because there was no ICU bed available. Investigation of the circumstances revealed that the ICU bed was occupied by patient C whose condition would allow transfer to the regular ward but the regular ward bed was occupied by patient D who was ready for discharge but was awaiting arrival of a family member to transport him to his home. Bed occupancy within the hospital had been at saturation for both ICU and regular ward beds for several weeks.

The high occupancy situation was managed by nurses and administrators by pairing new postoperative admissions with anticipated patient discharges, matching expected discharge and expected end of surgery times. Senior hospital management became involved in moment to moment decision making about bed allocation, surgical procedures starts, and intra-hospital patient transfers. Managers also sought increased efficiency of resource use, mainly through direct inquiries about patient status. A new administrative nursing position was established to centralize and rationalize bed resources. The system remained solid for approximately 5 weeks.

To avoid accidents, organizations struggle to keep the operating point away from the boundary of acceptable performance. Over time, this produces a marginal boundary that marks the acceptable limit of operations. Accident risk is deemed acceptably low as long as the operating point stays within this marginal boundary. Deliberate crossing of the marginal boundary violates the social norms of the organization. Uncertainty about the current location of the operating point and the dynamics of its movement may lead to unintentional crossing of the margin, but recognition that the operating point is outside the marginal boundary produces immediate efforts to bring it back inside the margin. These episodes are often described as “near miss” events because they cross the margin but do not cross the acceptable performance boundary. Real systems frequently operate near the marginal boundary so that workload and economic costs are minimized. Operations far away from the boundary tend to migrate down the gradients towards the marginal boundary because they are too expensive or too effortful to be sustained in the absence of accidents.⁶

In this model, safety is the product of the current operating point of the system, its propensity to move past the marginal boundary, and the relationship between the marginal and accident boundaries (fig 1B). Stable low risk systems (A) have operating points that move in small increments, remaining well away from the accident boundary. Stable high risk systems (B) operate nearer the margin and therefore nearer to accidents. In these systems, safety comes from the ability to keep the usual operating point movements small. Unstable systems (C) have sudden large operating point shifts that can take them from low to high risk.

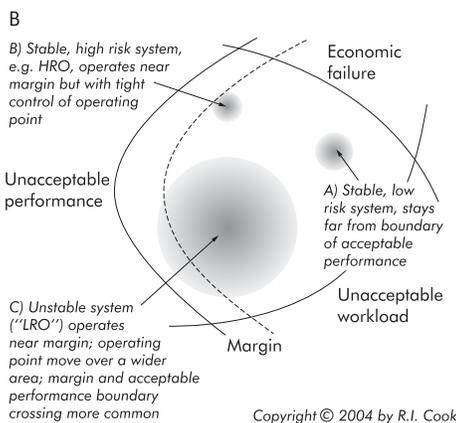
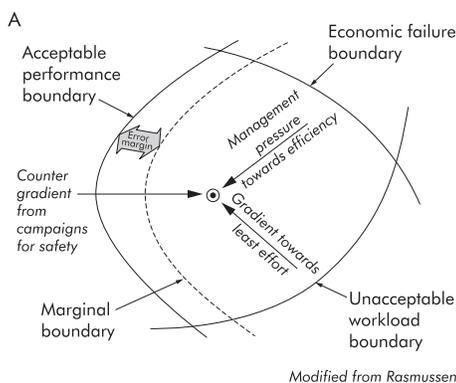


Figure 1 Dynamic safety model. (A) Gradients push the system operating point away from the boundaries of economic failure and work overload and towards the unacceptable performance (accident) boundary. (B) Stable low risk systems (A) operate far from this boundary; stable high risk systems (B) operate nearer the acceptable margin but the operating point moves in small increments and remains largely inside the marginal boundary; unstable systems (C) have large rapid shifts in the operating point. Modified from Rasmussen.⁵

The marginal boundary location is variable because it is controlled by sociotechnical processes. Publicized painful accidents usually lead to stepwise movement of the boundary inwards. Long periods without such events may result in marginal boundary creep outwards. The gradients encourage

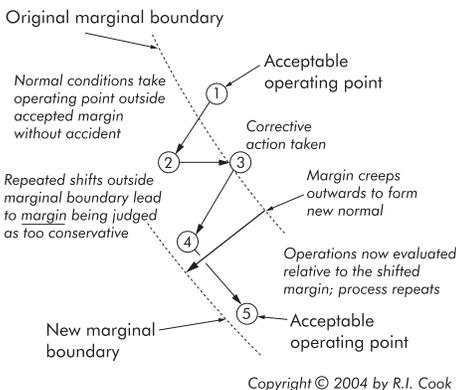


Figure 2 Flirting with the margin and marginal creep. The location of the marginal boundary is determined by sociotechnical processes. Over time, excursions of the operating point beyond the marginal boundary (“flirting with the margin”) that are not accompanied by overt accidents may lead to outward creep of the marginal boundary and regular operation of the system in what used to be the marginal region.

organizations to “test” the validity of the marginal boundary by deliberately moving the operating point beyond it. Because this “flirting with the margin” does not immediately produce accidents, it can lead to incremental adjustment of the marginal boundary outwards (fig 2).

Relatively fixed marginal boundaries and deliberately restricted operating point dynamics are the hallmark of so called “high reliability” organizations (HROs).⁷ HROs are capable of developing accurate, precise, and shared understanding of the current operating point location—factors that influence operating point movements and the distance between the point and the margin, as in (B). These organizations are usually targets of regulatory intervention and control—for example, the US commercial aviation or nuclear power industries. Low reliability organizations (LROs) are characterized by inaccurate, imprecise, divergent understandings of the current operating point and influential factors. Accidents in HROs are frequently anticipated while accidents in LROs are frequently astonishing “fundamental surprise” events.⁸

COUPLING AND ACCIDENTS

Tight coupling is a source of accidents in complex socio-technical systems.⁹ Tightly coupled systems have multiple, critical, concatenated dependencies that propagate effects through time and space, often via linkages that are difficult to anticipate. The increased complexity of tightly coupled systems tends to make it difficult to anticipate how such systems will behave under unusual conditions, to troubleshoot the systems during operation, and to analyze their behaviors after accidents. Loosely coupled systems are easier to analyze, which makes troubleshooting and new interventions easier.¹⁰

Tight coupling is usually introduced in order to gain efficiency and economies or to reduce the incidence of relatively low consequence events. These gains are obtained at the price of complexity and interdependences that lead accidents in such systems to be harder to foresee and prevent, larger in scale, and harder to recover from. Sudden large scale electrical power outages in the US and UK are convenient examples of tightly coupled system failures.

Although tight coupling occurs in some hospital settings,¹¹ hospital operations are usually loosely coupled. Individual units are independently staffed and have some degree of local autonomy. This arrangement produces operational “slack” that can buffer consequences of high workload in one unit. If an ICU is full, for example, it is possible to keep new critically ill patients in the emergency room or recovery room. This buffering capacity is lost, however, when the facility goes solid and *all* units are filled. Then even minor events in one unit may be major determinants of operations in the others.

“Going solid” changes the operational dynamics of the system but does not directly produce accidents. Instead, it makes accidents more likely, more difficult to foresee, harder to defend against, and harder to recover from. Without the simplicity and buffering that loose coupling provides, the system becomes brittle and difficult to manage. Normally mundane operational processes such as discharge of a patient from the hospital take on increased significance. Normally insignificant events such as a delay in discharge have increased consequences.

“Going solid” creates new work for practitioners and managers and new opportunities for failure. It increases the pressure to discharge patients, puts a premium on practitioner ability to identify candidates for discharge, and encourages them to speed the completion of care at each stage. The phenomenon involves the facility rather than a single unit. This leads managers to demand information about the conditions in each unit, further increasing the

performance pressure. The situation encourages high management to try to control usually routine details of daily operations, bringing senior management into direct contact with practitioners. The clinical conditions of patients take on organizational and managerial significance and management/practitioner conflicts are common. Even though it creates myriad difficulties, the economic payoff for going solid can be high. Because the fixed costs of hospital operations are spread over more patients, the hospital’s financial performance may improve, particularly if economic losses from overt failures can be avoided.

REACTIONS AND CONSEQUENCES

“Going solid” generates new performance demands. Clinical decision making becomes pressurized as practitioners struggle to identify the patients most likely to do well with lower levels of care in order to create openings for new patients. The average workload within units rises as the least sick are moved out to accommodate those who are more sick. The situation is ripe with perverse incentives for practitioners to “game” the larger system in order to manage workload. By overstating the acuity of patients in a unit, nursing workload measures may be manipulated in order to garner nursing resources. New patient admissions may be delayed by misrepresenting the bed availability in that unit—for example, by claiming the bed location is being “cleaned”. Some of the gaming can be quite subtle. A practitioner may “block” an ICU bed by keeping it occupied by a patient ready for discharge to the ward, timing the discharge of that patient in order to recapture the bed for another patient—for example, for an elective surgical procedure that will require an ICU bed for postoperative care. Paradoxically, although the incentives to hide resources in this way arise because the solid system has little available “slack”, these locally effective strategies perpetuate the solid condition.

“Going solid” is represented by an abrupt increase in the size of operating point motions (fig 3). The likelihood that the operating point will end up beyond the marginal boundary depends on the size of the movement and the original position of the operating point. Because “going solid” is likely to occur when the operating point is already near the marginal boundary, the transition to tight coupling may allow otherwise minor changes in the operating point to propel the system beyond the marginal boundary and produce an accident. This is especially likely if there has been substantial marginal boundary creep during a prolonged accident-free period.

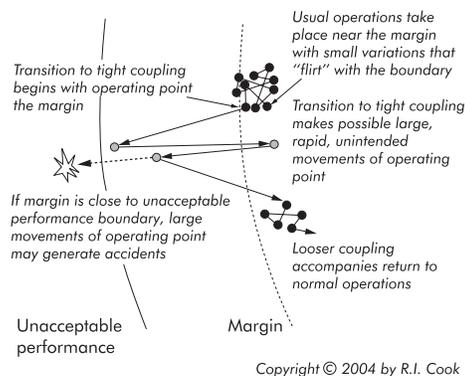


Figure 3 Consequences of “going solid”. Shifting from loosely to tightly coupled operations (going solid) makes possible large changes in the operating point location. This change in system dynamics may lead to movement of the operating point beyond the marginal boundary or even beyond the boundary of acceptable performance, resulting in an accident.

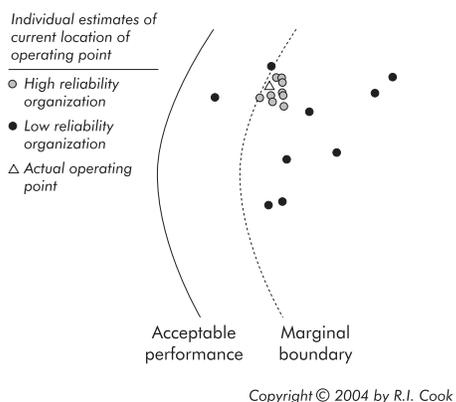


Figure 4 Characteristics of the safety culture of a high reliability organization (HRO). A necessary element of safety culture is a commonly held precise and accurate view of the current operating point of the system and its relationship to the margin and acceptable performance boundaries. Good quality tradeoffs are not possible if the location of the operating point is inaccurately known. Agreement about tradeoffs is unlikely if there are divergent views about the location.

IMPLICATIONS

The likelihood of accidents depends on how far the operating point is from the boundary of unacceptable performance and how far and how quickly the operating point can move. Using this model, a qualified understanding of safety includes a description of the point’s location, its distance from the marginal boundary, the location of the accident boundary relative to the marginal boundary, and of the dynamics of the point’s possible movements.

Current work on safety may be divided into three categories: (1) efforts to reduce the economic and workload gradients or to establish a counter gradient; (2) efforts to move the marginal boundary inwards; and (3) efforts to move the boundary of unacceptable performance outwards. At present there is little or no work directed towards characterizing the location or movement of the operating point of the system or reducing the size of operating point motions.

Although management has some influence over gradients and may seek to reduce them, especially immediately after local accidents, gradients are not under direct management control. Creating and sustaining counter gradients is difficult, especially when systems appear to be operating successfully. Infrequent experience with accidents is likely to result in “flirting with the margin” and marginal creep.

There is considerable interest in the “culture” of health care and discussions about what a safety culture is, how it can be produced and measured, and how to sustain it. These discussions can be grounded in model terms by understanding (a) the factors that determine the location of the marginal boundary relative to the boundary of acceptable performance, (b) the individual and organizational responses to approaching or crossing the marginal boundary, (c) the degree of consensus about the current location of the operating point relative to the boundaries, and (d) the accuracy of that consensus (fig 4). Within a system, divergent estimates of the current location or the potential movements of the point are a potent obstacle to thoughtful near term management of the operating point and to longer term work on safety. Resource hiding signals the presence of just such divergence. Efforts to relocate the marginal boundary are unlikely to be fruitful without agreement about the location of the operating point.

Current enthusiasm for information technology (IT) is based on the belief that IT applications will move the

Key messages

- According to Hirschhorn’s “law of systems”, every system moves to the limit of its capability.
- Safety is a dynamic property of healthcare systems.
- Tight coupling in systems leads to unexpected failures that are hard to control.
- Healthcare systems can become saturated with work and become tightly coupled.
- The dynamics of high reliability systems are well understood and carefully managed.
- Shared, accurate, and precise knowledge of the dynamics and location of the current operating point and boundary locations is a necessary component of a safety culture.

unacceptable performance boundary outwards. The marginal boundary is malleable, however, and these gains may be offset by marginal boundary creep. IT is itself a potent form of coupling¹⁰ and may generate large movements of the operating point. The tighter coupling may also limit practitioners’ ability to move the operating point away from the marginal boundary in response to local conditions, as is seen in some applications of bar coded medication administration.¹² New IT is introduced in high reliability settings such as commercial aviation with great care and deliberation. We believe that applications of IT in health care deserve similar critical assessment.

Finally, the willingness of organizations to tolerate going and remaining solid for long periods is likely to encourage flirting with the margin crossing and marginal creep. The ability to map the operating point of individual systems through time and study of the factors influencing the marginal boundary location are likely to be productive lines of inquiry. We note that there are few research results in this area.² Metrics capable of capturing “going solid” and its consequences would aid safety research work and public policy analysis.

CONCLUSIONS

In the past, loose coupling in hospitals buffered the short time scale fluctuations in workload. Managerial and technical changes have led to organization-wide improvements in efficiency. The result is the potential for complete work saturation (“going solid”) across the organization. “Going solid” creates new demands on workers and management and changes the properties of the healthcare system, setting up conditions conducive to accidents. The dynamic safety model presented here is a platform for investigating these systemic properties and its potential for creating accidents.

ACKNOWLEDGEMENTS

The authors acknowledge the generous contribution of multiple cases from physicians, nurses, pharmacists, technicians, and administrators.

Authors’ affiliations

R Cook, Department of Anesthesia and Critical Care, University of Chicago, Chicago, USA

J Rasmussen, Smorum Bygade 52, DK 2765, Denmark

Dr Cook is supported by the US Agency for Healthcare Research and Quality (AHRQ grants HS11816 and HS14261) and the National Library of Medicine (grant LM07947).

REFERENCES

- 1 **Woods DD**, Cook RI. Nine steps to move forward from error. *Cognition Technol Work* 2002;**4**:137–44.
- 2 **Cook RI**, Woods DD, Miller C. *A tale of two stories: contrasting views of patient safety*. Chicago, IL: NPSF, 1998. Available at <http://www.ctlab.org>.
- 3 **US Nuclear Regulatory Commission**. *25th Anniversary of Three Mile Island Unit 3 presentation*. Available at <http://www.nrc.gov/reading-rm/doc-collections/commission/tr/2004/20040303a.pdf>.
- 4 **Render ML**, Hirschhorn L. An irreplaceable safety culture. *Crit Care Clin* 2005;**21**:31–41.
- 5 **Rasmussen J**. Risk management in a dynamic society: a modeling problem. *Saf Sci* 1997;**27**:183–213.
- 6 **Vaughn D**. *The Challenger launch decision: risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press, 1997.
- 7 **Rochlin G**, LaPorte T, Roberts K. The self-designing high reliability organization: naval carrier operations at sea. *Naval War College Review* 1987;Autumn:76–90.
- 8 **Lanier Z**. *Fundamental surprise: the national intelligence crisis*. Tel Aviv: Hakibutz HaMeuhad and Jaffee Center, 1983.
- 9 **Perrow C**. *Normal accidents: living with high risk technologies*. Princeton: Princeton University Press, 1984.
- 10 **Cook RI**, Woods DD. Operating at the sharp end: the complexity of human error. In: Bogner S, ed, *Human error in medicine*. Hillsdale, NJ: Lawrence Erlbaum, 1994:255–310.
- 11 **Gaba D**, Maxwell M, DeAnda A. Anesthetic mishaps: breaking the chain of accident evolution. *Anesthesiology* 1987;**66**:670–6.
- 12 **Patterson ES**, Cook RI, Render ML. Improving patient safety by identifying side effects from introducing bar codes medication administration. *J Am Med Inform Assoc* 2002;**9**:540–53.