Modeling Rasmussen's dynamic modeling problem: drift towards a boundary of safety

J. Bradley Morrison¹ · Robert L. Wears²

Received: 30 June 2018 / Accepted: 18 January 2021 / Published online: 28 April 2021
© The Author(s) 2021

Abstract

We build a system dynamics model based on a conceptual model originally proposed by safety scientist Jens Rasmussen to explore the dynamics of a safety system subject to pressures for performance improvement. Rasmussen described forces that generate a drift in the boundary of acceptable performance that can push the organization towards “flirting with the margin” and thus operate at very high risk of catastrophic safety failure. Simulations of the model faithfully replicate the behavior described by Rasmussen and others in a variety of scenarios. Simulation experiments further illuminate the potential for risky behavior and point towards some approaches to better system safety.

Keywords Safety science · Human factors · System dynamics · Organizational drift · Dynamic modeling

1 Introduction

The influence on organizational performance of various organizational or individual objectives that are not entirely aligned is a common theme that appears in many streams of the scholarly and practitioner literature. Analyses of safety breaches as witnessed in tragic examples such as the space shuttles Challenger and Columbia (Starbuck and Farjoun 2005; Vaughan 1996), petroleum industry disasters at Texas City and Deepwater Horizon (Hopkins 2008, 2018), nuclear power accidents at Three Mile Island and Chernobyl (Perrow 1984), and coal mining explosions at Massey's Upper Big Branch mine and at New Zealand's Pike River mine (Galuska 2012) call our attention to the devastating consequences that can result when objectives for safe operations yield to pressures to achieve other goals such as deadlines and budgets. In this paper, we use a system dynamics model based on a conceptual model originally proposed by the influential safety scientist Jens Rasmussen to explore the dynamics of a safety system subject to pressures for performance improvement.

The interplay of competing organizational priorities is a central theme in the study of a variety of organizational phenomena (Cyert and March 1963; Nelson and Winter 1982). Safety failure cycles and organizational drift into failure are well-documented modes of behavior theorized to result from such competing priorities (Dekker 2016; Rasmussen 1997; Reason 1997). Organizational pressures for productivity exert influence on day-to-day behavior and can lead to a gradual erosion of safety standards, moving the organization closer to the boundary of safety failure. Efforts to improve productivity often yield immediate, tangible, and visible gains with a high degree of certainty, whereas efforts to maintain safety generate feedback that presents as ambiguous, intermittent, and often delayed and uncertain (Farjoun 2005). Reliability is “invisible” so there is “nothing to pay attention to” (Weick 1987). Consequently, organizations find it challenging to maintain a focus on safety and to continue to adhere to safety standards for extended periods of time, particularly in the absence of safety breaches that provide powerful reminders of the riskiness of the organization's activity. For example, analyses of the predecessors of the *Columbia* disaster reveal the erosion of technical requirements and safety procedures in the decade leading up to the disaster (Farjoun 2005; Woods 2005). Defenses gradually erode in the face of production pressure, so the dynamic over time is a drift towards failure.

To extend our understanding of organizational drift, in this paper we examine an existing theory, formalizing it to

✉ J. Bradley Morrison
bmorriso@brandeis.edu

¹ Brandeis University, Waltham, MA, USA

² University of Florida, Jacksonville, FL, USA

assess how well the theory accounts for the phenomena the theorist aimed to explain. Our focus is Rasmussen's (1997) theory of migration towards the boundary of acceptable performance, in which normal changes in daily work conditions give rise to frequent modifications in strategies and activities. These situation-induced variations allow for actors to discover an "effort gradient" that coexists with an "efficiency gradient" (often supplied by managers). The interactions can very likely result in "a systematic migration towards the boundary of functionally acceptable performance and, if crossing the boundary is irreversible, an error or accident may occur" (Rasmussen 1997: 189). Because he presents an explicit theory (comprising the variations and the gradients), displays the logic of the theory in diagrammatic form, and describes the dynamic patterns of behavior (migration towards the boundary) it explains, the causal argument Rasmussen presents is an ideal candidate for a detailed and systematic examination. Our purpose is to examine the consistency, completeness and parsimony of the causal explanation contained in the existing theoretical model.

Modeling organizational pressures, tradeoffs, and decisions follows a rich tradition in the study of organizations more broadly, not just in safety science (Cyert and March 1963; Nelson and Winter 1974), and such formalisms can lead to new insights.

Peli and colleagues present one approach to such theory examination, using first-order logic to analyze Hannan and Freeman's theory of organizational inertia (Hannan and Freeman 1984; Peli et al. 1994). An exemplar precedent for the approach we follow is found in Sastry's formalization of Tushman and Romanelli's (1985) theory of punctuated equilibrium (Sastry 1997; Tushman and Romanelli 1985). Sastry developed a system dynamics model of the theory, drawing directly from the textual analysis of the 1985 paper. As we do here, she chose system dynamics because a causal modeling approach suitable for capturing dynamics is needed. We use system dynamics as a simulation technique to "provide an analytically precise means of specifying the assumptions and theoretical logic that lie at the heart of verbal theories" and offer "superior insight into complex theoretical relationships among constructs" (Davis et al. 2007: 480). Following Sastry, we use the method to formalize verbal descriptions of causal relations and test the theory's ability to explain the dynamic phenomena described by the authors.

2 Context for Rasmussen

Safety as a topic arises in a range of scholarly fields such as human factors, safety science, human error, accident research, organizational studies, operations management, engineering, psychology, sociology, and system dynamics. Jens Rasmussen, one of the most influential contributors

to this scholarly discourse, published a paper in *Safety Science* in 1997 entitled "Risk management in a dynamic society: a modelling problem". The model, presented using a diagram and a verbal explanation, was offered to explain the pattern of behavior that resulted in an organization's operating practices drifting towards, and possibly crossing, a boundary of safe behavior (Rasmussen 1997).

Before turning to a description of Rasmussen's model, we provide some context by considering how ideas about accidents, risk, and safety have evolved over time.

At the time of Rasmussen's model, thinking about safety had been shifting. For much of the last century, scholars of safety conceived of accidents as having some special cause, different from what caused ordinary success. Popular models in this tradition included a domino model (Heinrich 1931) a Swiss cheese model (Reason 1997) failure modes and effects analysis (Ben-Daya and Raouf 1996), models of human error due to cognitive limitations (Lopes 1986; Tversky and Kahneman 1974), and a variety of models in healthcare that focused on human fallibility as a cause of medical error (Hurwitz and Sheikh 2009). In the wake of several highly visible accidents such as the 1977 Tenerife disaster and the 1979 accident at Three Mile Island, thinking was gradually shifting towards views of accidents as emerging from the interaction of a multitude of events, processes and relationships in a complex system (Rasmussen et al. 1990; Woods et al. 2010: 36). Scholars began to question the construct of human error as an external object and adopted a view of human error as a social construction (Hollnagel 1983, 1991; Rasmussen et al. 1987; Rasmussen and Batstone 1989).

Rasmussen went further in arguing that the self-organizing and adaptive features of complex systems will necessarily lead to accidents because the actions of individual workers constantly adapting within local constraints can sometimes combine to cause catastrophic failures. Human performance hinged on continuously making sense of a jumbled stream of phenomena to maintain cognitive control (Klein et al. 1993; Reason 1990). In a similar vein, other scholars viewed accidents as emerging from "organizational drift", as safety standards and generally accepted procedures drift over time in response to perceived risks, economic pressures, and workload issues (Dekker 2011; Snook 2000).

Rasmussen's ideas have been described as a standard of thinking in cognitive engineering (Reason 1990: xiii), a paradigm shift (Moray 1988: 12), and the most important body of thought shaping modern safety science (Wears 2017). Rasmussen's work contributed important conceptual tools, such as the abstraction hierarchy (Rasmussen 1985), the skills-rules-knowledge framework (Rasmussen 1983), the role of error in organizing behavior (Rasmussen 1990), the problems of causally attributing accidents to errors (Rasmussen et al. 1990), and the dynamic nature of

problem-solving strategies (Rasmussen 1993) and of safety itself (Rasmussen 1997).

Rasmussen's Dynamic Safety has special importance in the realm of safety science because of the combination of two characteristics that made it path-making. First, it reflects a view of safety as a property of a system not as the absence of human error. It focuses attention on the structure and behavior of the system, not on the actions of individuals to be blamed for accidents. Second, it is a compact, causal theory that describes dynamics and proposes a system structure that generates those dynamics. It lacks the formalism of a rigorously specified system dynamics model, but it contains the causal logic typically seen in conceptual models that precede such formal models (Randers 1980).

The model development and analysis that follow are motivated by Rasmussen's Dynamic Safety conceptual model. In the next section, we briefly review Rasmussen's model. We then describe our research approach. In the following two sections, we present the system dynamics model we built to examine the dynamics. In the following section, we conduct a series of experiments with this model to explore its behavior with an aim to replicate the verbal results described by Rasmussen and other commentators. In the final section, we discuss the results and suggest some directions for further work.

3 Rasmussen's dynamic safety model

Rasmussen (1997) began with the notion that the behavior of people in a system "is shaped by objectives and constraints which must be respected by the actors for work performance to be successful" (p. 189). Their work "is bounded by administrative, functional, and safety related constraints" (p.189). "Frequent modifications of strategies and activity will show great variability" that Rasmussen likened to 'Brownian movement' of molecules of a gas (p.189). Workers will identify 'an effort gradient' and management will normally supply an effective 'cost gradient'. "The result will very likely be a systematic migration toward the boundary of functionally acceptable performance and, if crossing the boundary is irreversible, an error or an accident may occur" (p. 189). Rasmussen's depiction of this conceptual model is shown in Fig. 1.

In a later paper, Cook and Rasmussen (2005) described the model further:

"Rasmussen's dynamic safety model ... describes the feasible operating space for a sociotechnical system within three boundaries that form an envelope. 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

Fig. 1 Rasmussen's dynamic safety model as depicted in Cook and Rasmussen (2005)

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 the distance from the unacceptable performance boundary increases. In practice, the precise location of the boundary of unacceptable performance is uncertain." (p. 130).

4 Research approach

As Rasmussen's title suggests, managing the risk rooted in a dynamic society presents a modeling problem. A causal, dynamic model that is capable of simulating the behavior

that evolves from the interrelating variables serves as an ideal tool for analyzing organizational drift. Because Rasmussen's theory is a theory of dynamics—that is, of how things change over time—the modeling and simulation capabilities of system dynamics seem especially well suited for such formalizing. System dynamics (Forrester 1961; Sterman 2000) provides a simulation method that effectively represents human activity and change processes and that has already proven useful for testing many theories rooted in sociological, economic, socio-technical, social psychology, and individual psychology domains. Notable examples using system dynamics to examine and apply existing organizational theories include examinations of a punctuated equilibrium model of organizational change (Sastry 1997), a theory of the co-evolution of technology and organization (Black et al. 2004), an explanation for drift in compliance with regulations and rules (Martinez-Moyano et al. 2014), a feedback theory of process improvement failure (Morrison 2012), and a study of the Yerkes-Dodson curvilinear relationship of performance to stress applied to understanding the evolution of disasters (Rudolph and Repenning 2002; Yerkes and Dodson 1908). Simulation techniques have a rich history of scholarly studies to examine organizational behavior (Levinthal 1997; March 1991) and are becoming an increasingly significant approach to theory development (Davis et al. 2007; Harrison et al. 2007). System dynamics differs from other formal modeling techniques (Axelrod 2007; Carley and Wallace 2001) in several important ways.

First, system dynamics features feedback processes describing circular causal relationships in which variables both influence and depend on each other (Richardson 1991). The feedback perspective emerged from cybernetics and engineering control theory in the mid twentieth century and has since then been used to understand a variety of social processes, such as self-fulfilling prophecies (Merton 1948), vicious cycles (Masuch 1985), organizational demise (Hall 1976), process improvement programs (Sterman et al. 1997), action-oriented problem solving (Rudolph et al. 2009), and organizational learning (Rahmandad 2008). System dynamics has also been applied in the safety domain to study mine disasters (Cooke 2003), construction accidents (Mitropoulos et al. 2005), incident learning systems (Cooke and Rohleder 2006), combat vehicle accidents (Minami and Madnick 2009), construction safety culture (Mohamed and Chinda 2011), traffic safety policy (Goh and Love 2012), chemical storage (Bouloiz et al. 2013), construction worker attitudes (Shin et al. 2014), coal mine safety (Liu et al. 2015), and others. System dynamics rests on exactly the perspective that Rasmussen himself called for: “an active, closed loop feedback point of view” (Rasmussen 1997: 192).

Second, system dynamics incorporates explicit representations of behavioral decision-making and the pressures that influence actors (Morecroft 1985; Morrison and Oliva 2018; Sterman 1994). A system dynamics model integrates the operational and behavioral aspects of a system into a formal representation of the theory of dynamics. The system dynamics feedback model represents the physical and institutional structure of the system coupled with the decision rules representing the behavior of various actors in the system (Sterman et al. 2015). Modelers strive to represent the various influences and processes to achieve a high degree of operational character, carefully describing the causal pathways that affect changes in the variables describing the system (Richmond 1993). By explicitly representing the processes that govern change with operational descriptions of feedback processes, the model enables simulation of the dynamic *behavior* that results from the *structure* described in the model. The fundamental premise of system dynamics is that the behavior of a system arises from its endogenous structure (Richardson 2011). Thus, simulations test the internal consistency and coherence of a theory and its ability to explain dynamic behavior.

Third, system dynamics distinguish between stocks or state variables, such as the marginal boundary and the operating point, and variables that represent rates of change, such as the rate of change in the marginal boundary and adjustments to the operating point. Distinguishing these types of variables is important, because the stocks characterize the system. They represent properties of the system that have accumulated over time and importantly cannot be changed instantly. They hold the system's “memory” of past adjustments and as such they give the system its dynamics. Changes in the system—the decisions and actions of actors—are based on information about these state variables or stocks. System dynamics models describe continuous-time processes rather than discrete processes and therefore model ongoing processes and simultaneous procedures that influence each other. The models can be used to explore the sensitivity of dynamic behavior to changes in model assumptions and parameter values including time lags.

Identifying constructs

The modeling strategy is to begin with the conceptual model Rasmussen offered; identify the concepts, relationships, and causal logic in his model, and translate those concepts and relationships into a mathematical representation of the feedback processes implied in his model. The goal was first to stay true to Rasmussen's original description, adding only what is essential to formalize his existing theory rather than refining any aspects of the theory as originally presented. The reasoning behind this disciplined adherence to the original is so that the resulting mathematical model can be used as a laboratory to explore the consequences of the *original*

theory. Subsequent iterations could be used to refine or extend the theory.

The first step in developing the formal representation of Rasmussen's theory was to identify constructs and relationships. We conducted a textual analysis of Rasmussen's 1997 paper and also of Cook and Rasmussen's 2005 sequel, extracting statements relevant for the model formalization and coding them into categories. Rasmussen's model focuses on the interaction between the forces in Fig. 1 denoted "counter gradient from campaigns for safety" and the other forces, namely the "management pressure towards efficiency" and the "gradient towards least effort." His concern is the possible migration of the operating point (seen not labeled in the center of Fig. 1) and possible excursions beyond the acceptable performance boundary (i.e., accidents). We identified statements describing constructs, collecting into categories those that appear to refer to the same construct, and we analyzed statements describing relationships between the constructs. We also identified qualitative descriptions of the dynamic patterns of behavior for later use in testing the model. Table 1 presents examples of statements that provide definitions of variables, assertions about causal relationships, and predictions of the dynamic behavior of the system.

Table 2 presents the four stock variables—marginal boundary, economic failure boundary, operating point Kernel, and perceived operating point—that emerged as the starting point for formalizing the causal framework of Rasmussen's theory.

5 Representing the causal structure of the theory

The next step in formalizing the model was to describe the relationships of the variables to each other to capture the causal mechanisms hypothesized to explain the dynamics. The result was a set of interrelated feedback loops that represent the pressures that give rise to the dynamics of organizational drift. Figure 2 depicts a summary of these relationships in a causal loop diagram following the standard convention of using an arrow between two variables to indicate the influence of one variable on another.

The feedback loops in Fig. 2 represent the processes that Rasmussen proposes to explain the movement towards the Acceptable Performance Boundary (i.e., the movement towards increased risk of accidents). To read Fig. 2, consider first a core construct in Rasmussen's theory: the operating point, describing the location in Fig. 1 that characterizes the organization's current operations. The theory suggests that the operating point moves continuously and with great variability as a result of adaptive search guided by socio-technical processes. In Fig. 2, the operating point

Table 1 Summary of textual coding categories

Coding category	Definition	Structure/relationship		Dynamic behavior	
		Explanation	Example	Description of how one variable influences another	Pattern of variables' evolution over time
Explanation Example	The work space within which the human actors can navigate freely ... is bounded by administrative, functional and safety related constraints." (Ras: 189) The "Resulting perceived boundary of acceptable performance" is separated from the "Boundary of functionally acceptable performance" by a distance labeled the "Error margin." (Ras: 190, Fig. 3) "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 the dynamics of the points possible movements." (C&R: 133)	"The boundary of safe behavior of one particular actor depends on the possible violation of safe defenses by other actors" (Ras: 190) "Degrees of freedom will have to be closed by an individual actor guided by adaptive search guided by process criteria such as work load, cost effectiveness, risk of failure ..." (Ras: 189) "Changes in the gradients (for example, increased economic pressure) move the operating point	"Natural migration of activities toward the boundary of acceptable performance" (Ras: 189) "If crossing the boundary is irreversible an error or accident may occur" (Ras: 189) "A quite normal variation in somebody's behavior can then release an accident" (Ras: 190) "Rasmussen's dynamic model ... can be used to formulate a ... safety model that includes 'going solid' and its consequences." (C&R: 130)	"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." (C&R: 132)	

Ras refers to Rasmussen (1997). C&R refers to Cook and Rasmussen (2005)

Table 2 Stock variables in the Rasmussen gradient model

Construct	Related constructs	Description
Marginal boundary	Objectives and constraints (Ras:189); occupational risk, major accidents (Ras:189); Boundary conditions of the workspace (Ras:191)	“Over time, this produces a marginal boundary that marks the limits of acceptable performance.” (C&R: 131)
Economic failure boundary	Objectives and constraints (Ras:189); administrative constraints (Ras:189); cost gradient (Ras:189); Pressure towards cost-effectiveness (Ras:190); Boundary conditions of the workspace (Ras:191)	“feasible operating space ... within three boundaries that form an envelope” (C&R: 130)
Operating point Kernel	Work performance (Ras:189) “The operating point location is influenced by gradients that drive operations away from the workload and economic failure boundaries...” (C&R: 130)	Historical point from which “situation-induced variations within the work space” occur (Ras: 189)
Perceived operating point	“HROS are capable of developing accurate, precise, and shared understanding of the current operating point location ... LROs are characterized by inaccurate, imprecise, divergent understandings of the current operating point.” (C&R: 132)	Understanding of the current operating point location (C&R:132)

Ras refers to Rasmussen (1997). C&R refers to Cook and Rasmussen (2005)

actual results from the combination of excursions with the recent operating point, named the operating point kernel. The arrows in the figure describe causal relationships indicating that a change in the first variable (at the tail) causes a change in the second variable (at the arrowhead). A “+” sign indicates that a change in the first variable causes a change in the second variable in the same direction. A “−” sign indicates that a change in the first variable causes a change in the second variable in the opposite direction. We define moving down or decreasing the location of the operating point as moving towards a greater risk of accidents, that is, towards the boundary of acceptable performance.

Rasmussen describes two forces that push the system towards this boundary: “management pressure towards efficiency” and the “gradient towards least effort,” as well as a “safety gradient” that pushes the system away from this boundary. Feedback loop B1a describes the effects of the efficiency gradient. Starting from the variable Operating Point actual and reading clockwise around the loop, we can see the process described. As the Operating Point Actual moves up (towards greater inefficiency, i.e., higher cost), the Efficiency Gradient Pressure increases, which in turn pushes the Operating Point Kernel downward. When Excursions combine with the lower Operating Point Kernel, the result is a lower Operating Point Actual (because the + sign means the second variable moves in the same direction, in this case lower). The process forms a feedback loop known as a Balancing Loop, because the process in the loop balances out, or offsets, the original change. Balancing loops bring stability to systems as they act to move the system towards their implied or explicit goals (Sterman 2000). The loop is acting to move the operating point in the direction of the efficiency gradient (and thus towards the boundary of acceptable performance), exactly as Rasmussen described.

Similarly, on the left of Fig. 2, we see balancing feedback loop B1b that describes the effects of the least effort gradient. It too is a balancing loop, and it acts to move the operating point in the direction of least effort, which is also towards the boundary of acceptable performance (downward in our representation). The right side of the diagram displays the feedback loop B2 that describes the effect of the safety gradient. Note the difference in the polarity of the arrows. As the Operating Point Actual decreases (moving towards less safety), the Safety Gradient Pressure increases, and this increase in pressure moves the Operating Point Kernel upwards (in the direction of greater safety). Thus, these three balancing feedback loops describe the processes through which the boundaries or constraints exert their influence to form an envelope within which the human activity takes place.

The core idea of Rasmussen’s theory is his explanation for the “natural migration of activities toward the boundary of acceptable performance” (Rasmussen 1997:

Fig. 2 Causal loop diagram of Rasmussen's 1997 theory. Arrows represent causal relationships. An arrow with a+ sign indicates that an increase (decrease) in the variable at the tail of the arrow causes an increase (decrease) in the variable at the head of the arrow. That is, the + sign indicates the consequent variable moves in the same direc-

tion. An arrow with a- sign indicates that an increase (decrease) in the variable at the tail of the arrow causes a decrease (increase) in the variable at the head of the arrow. That is, the - sign indicates the consequent variable moves in the opposite direction.

189). The logic of change as he describes it can be seen in the Loop R3 in Fig. 2. Encouraged by the gradients for efficiency and least effort, Excursions move the Operating Point Actual downward (towards less safety). These Excursions do not immediately lead to accidents, instead sending the signal that the location of the operating point is not problematic. This experience leads to an incremental adjustment of the marginal boundary in the downward direction. With a lower marginal boundary comes a reduction in the safety gradient pressure, as the operating point actual is now further from the marginal boundary. With less safety gradient pressure, the operating point Kernel—the new normal operating point—also adjusts downwards. Future excursions from this new, lower operating point Kernel bring the operating point actually even lower, and the cycle continues. The process forms a feedback loop known as a reinforcing loop, labeled R3, because the process in the loop reinforces, or continues the direction of, the original change. Reinforcing loops bring instability to systems, and indeed this reinforcing loop is the heart of Rasmussen's theory (Sterman 2000). Familiar reinforcing loops include the contagious spread of an infectious disease, the “going viral” of a social media post, the growth of an organism by cell division, and the compounding of interest in the bank.

6 Formalizing Rasmussen's model

To enable ready analysis of the dynamic behavior of the operating point with respect to the acceptable performance boundary, we focus on the dimension that represents the distance from the acceptable performance boundary. To model the effects of the cost gradient and the effort gradient on movement along our focal dimension, we simply decompose the gradients into two orthogonal components. We consider the component vector of pressure arising from the management pressure towards efficiency that is in direct opposition to the counter gradient from campaigns for safety (e.g., along the x axis of Fig. 1) added to the component vector of pressure arising from the gradient towards least effort that is in direct opposition to the counter gradient from campaigns for safety to form one aggregate pressure. This composite vector of economic and least effort pressure will push the operating point in the direction towards the boundary of acceptable performance (e.g., along the x axis). The other component vectors of these gradients (e.g., along the y axis) would determine the location of the operating point between the economic failure boundary and the unacceptable workload boundary (e.g., along the y axis). The location along this dimension reflects the relative dominance of managerial pressures compared to worker pressures, but it does not inform the immediate questions regarding risk and safety. Given our focus on safety and the location of the operating point relative to the boundary of acceptable performance

Fig. 3 Constructs in the dynamic safety model

and the marginal boundary, it will be sufficient to represent distance in one dimension (e.g., the x axis). Doing so allows us to conveniently represent a time series of behavior using a two-dimensional plot of distance and time. Note, we can with no loss of generality limit our consideration to one dimension. A model showing two dimensions of distance could also show migration along the orthogonal dimension, but it would show the same behavior along our focal dimension.

Figure 3 shows the relationship of key ideas in Rasmussen's model and our representation of it as simplified in one dimension. The focal dimension is the vertical or y axis in Fig. 3. On one edge (shown at the bottom) of the spectrum of performance lies the acceptable performance boundary. Excursions below this boundary are accidents, safety breaches, or safety failures. The distance from the true boundary of failure and the marginal boundary constitutes an error margin. At the opposite edge of the performance spectrum lies the economic failure boundary.

In the remainder of this section, we describe how we formalized the feedback structure depicted in Fig. 2. In moving from a causal loop diagram to a more formal, mathematical description of the theory, we follow standard model formulation practices in system dynamics to represent the causal logic implicit in the arrows shown in Fig. 2. Each arrow represents a causal influence, so the equation for each variable in the diagram is a function of the variables with arrows that point into the focal variable. The result is a model that translates the verbal theory into a set of equations that define the endogenous relationships comprised by the theory as described by Rasmussen. The causal structure is defined within the model, so the model can be used to test and explore the behavior of the set of assumptions or causal

Fig. 4 Stocks representing the landscape in the dynamic safety model

assertions that form the theory. Here we describe the translation of Rasmussen's causal logic into the model, omitting the mathematical equations. The fully documented model including equations is provided as a supplementary file in support of this paper.

The first key idea in Rasmussen's model is the definition of the landscape based on the several boundaries surrounding the operating point. Figure 4 displays the stock variables (See Table 2) used to represent the landscape depicted in Figs. 1 and 3. Starting from the bottom, we see the variable

Acceptable Performance Boundary. The Marginal Boundary is depicted as a rectangle signaling that we model this construct as a stock variable. A stock models an accumulation process; that is, its value is the accumulation of the increases to it (known as inflows) and decreases to it (known as outflows) over time. Rasmussen asserts that this Marginal Boundary can change over time, moving closer to or further from the acceptable performance boundary. The variable marginal boundary will represent the location as a distance from the acceptable performance boundary (arbitrarily set as the zero point for measurements in our model). The error margin is the distance between the marginal boundary and the acceptable performance boundary.

At the top of Fig. 3, we see the economic failure boundary, also represented as a stock variable with a rectangle. The third rectangle, at the center of the diagram, is called the operating point kernel. Consistent with Rasmussen's description, we represent the operating point actual as the combination of an underlying operating point kernel that changes slowly over time and an excursion that is a deviation from the operating point kernel.

The second key idea in Rasmussen's model is that the boundaries (except for the acceptable performance boundary) are not fixed but may move over time in response to organizational experience and current pressures. Figure 5 diagrams the constructs used to model the changes to the location of the four stocks shown in Fig. 4. In each case,

there is a flow variable (represented with a pipe and valve icon) that changes the stock variable. For example, when the change in marginal value is positive, the marginal boundary increases, and when the change in the marginal boundary is negative, the marginal boundary decreases. These stock and flow representations follow standard practice and model the gradual adjustment of the boundaries and operating point kernel towards an implied target value, called the indicated value. For example, the flow change in economic boundary adjusts the stock economic failure boundary towards the indicated economic boundary. The direction and magnitude of change depend on the gap between the indicated and actual values and a time constant that defines the fraction of the gap closed each unit of time (weeks, in this case). Thus, stocks do not change instantly but only more slowly, representing the gradual migrations based on the conditions the system experiences, as described by Rasmussen. In the absence of other changes, the stock will eventually equal the indicated point.

The third key idea in Rasmussen's model is that there are gradients that exert pressure on the operating point. Figure 3 shows the two focal gradients, the counter gradient from campaigns for safety and the pressure towards efficiency. These are opposing forces that exert pressure along the dimension represented by the y axis in Fig. 3. Our modeling goal was to represent the pressure along this gradient, but to do so, we found that we needed to add an additional

Fig. 5 Flows to adjust stocks in the dynamic safety model. Pipe and valve icons signal flow variables. Flows when positive increase stocks and when negative decrease stocks. The direction and magnitudes of the flow depend on the differences between the stocks and their indicated points.

construct to fully specify the theory. It is not unusual when generating an explicit mathematical formalization of a verbal theory to discover some gaps in the explanation of the causal logic (Peli et al. 1994; Sastry 1997). In this case, Rasmussen clearly described the *direction* of the pressure from the safety gradient as directly away from the boundary of acceptable performance. However, the verbal theory is silent regarding the *magnitude* of the pressure from the safety gradient. The gap we identified is not simply a question of the actual value of a parameter defining the pressure but rather that there is a missing construct. In seeking to fill this gap, we sought the most parsimonious solution that would be consistent with Rasmussen's theory. The solution that emerged was to define a variable called safety goal that describes the desired location of the operating point implied by the gradient pressure for safety. This is a conceptually satisfying enhancement because the "goal" of the safety gradient pressure is conceptually distinct from the marginal boundary itself. The "goal" of the safety gradient pressure is to stay within—not on—the marginal boundary (except in the special, possibly degenerate, case where the safety goal actually equals the marginal boundary, a case of a rather weak gradient for safety). In Fig. 6, we have added to the diagram the operationalization of the gradient for campaigns for safety. Safety gradient pressure is rooted in a comparison of the perceived operating point to the safety goal. When the perceived operating point is far from the safety goal, the pressure will be minimal, but when the perceived operating point approaches the safety goal, safety gradient pressure

will be large so as to push the system away from the marginal boundary and towards greater safety. The perceived operating point is also a stock variable that adjusts toward the operating point actual. Occasional variations in the operating point will have a small effect on the perceived operating point, but sustained variations will result in greater changes to the perceived operating point. An increase in the safety gradient pressure causes an increase in the net pressure, which in turn increases the operating point and thus eventually the operating point Kernel. (Increases in the operating point in our landscape signify movements further from the acceptable performance boundary and are thus movements towards less risky operating points.)

Rasmussen's model also includes gradients that work in opposition to the Safety Gradient Pressure. His model includes both "management pressure towards efficiency" (arising from management or market forces for productivity or cost) and a "gradient towards least effort" (arising from worker behavior). Here, we represent the opposition to the safety gradient as the efficiency gradient pressure, recalling that this gradient comprises pressures arising from both forces, consistent with our choice to model the landscape in one dimension. Figure 7 completes our model by adding structure for the efficiency gradient pressure, exactly analogous to the safety gradient pressure. There is an implicit efficiency goal at a given distance from the economic failure boundary, where that distance is specified by the parameter desired distance from econ boundary. Comparison of the perceived operating point to the efficiency goal yields an efficiency gradient

Fig. 6 Safety gradient pressure in the dynamic safety model

Fig. 7 Feedback structure of the dynamic safety model

pressure, similar but opposite to the safety gradient pressure. When the efficiency goal is below the perceived operating point (i.e., the management desires more efficiency or lower costs), the efficiency gradient pressure acts as a multiplier of the operating point kernel to determine the indicated operating point. The two pressures combine to form net pressure, which exerts its influence on the indicated operating point and over time causes movement in the operating point kernel.

The model includes some functionality for testing and analysis purposes. The model also includes another inflow, not shown in the diagram, to the stock marginal boundary. The inflow is called crisis reset. It is triggered in the instance of a safety failure, defined as the operating point actual moving beyond the acceptable performance boundary. In such a case, the crisis reset inflow resets the value of marginal boundary to the initial value of the marginal boundary.

6.1 Rasmussen's theory reflected in the model

This section summarizes the mapping of Rasmussen’s dynamic safety model with the formalized model presented here. The causal theory in Rasmussen’s model asserts two causal mechanisms that combine to push the system towards “flirting with the margin” (Cook and Rasmussen 2005, p. 132). The first of these mechanisms is the pressure from management (and workers) to achieve

greater efficiency (less efforts) that causes a migration down the gradient away from the boundary of economic failure (unacceptable workload) and towards the marginal (safety) boundary. The balancing loops B1a and B1b in Fig. 3 operationalize the causal effects of this pressure, and in Fig. 7 we incorporate an Efficiency Goal to avoid getting too close to economic demise.

The second mechanism asserted in Rasmussen's theory is the effect of "Experiments to improve performance that create Brownian motion" (Rasmussen 1997, p. 190). Our model includes this mechanism in the determination of the operating point actual. The operating point actual is determined by the stock operating point kernel adjusted by the amount of excursion. The excursion is a random variable determined by pink noise, the variability of which can be adjusted by varying the parameter standard deviation.

Rasmussen also suggests there will be a pressure gradient arising from campaigns for safety. Our model includes a safety goal that is a given distance from the marginal boundary, where that distance is specified by the parameter *safe distance from the boundary*. Safety gradient pressure is the ratio of the safety goal to the perceived operating point. When the safety goal is above the perceived operating point (i.e., there is pressure to operate further away from the marginal boundary), the safety gradient pressure combines with the efficiency gradient pressure as a

multiplier of the operating point kernel to determine the indicated operating point. The mechanism of the counter-gradients from concern for safety is thus operationalized in balancing loop B2 in Fig. 3.

Finally, Rasmussen's theory claims that organizations will adapt to changing conditions. The locations of the boundaries will not change instantly, but they are variable due to their dependence on sociotechnical processes (Cook and Rasmussen 2005). Our model includes these adaptations using stock variables that adjust slowly over time based on the various pressures and the experiences (from the operating point actual) of the organization. By using the stock, flow and feedback representation, we heed Rasmussen's call: "Modeling risk management in a dynamic society in which all actors continuously strive to adapt to changes and the pressures of dynamic markets, we clearly have to apply such an active, closed-loop feedback point of view" (Rasmussen 1997: p. 192).

7 Dynamic behavior of the safety system

In this section, we present the results of experiments to explore the behavior of the safety system represented by the system dynamics model. In our first set of experiments, we test the dynamic hypotheses from Cook and Rasmussen (2005) by attempting to replicate the patterns of behavior described verbally in their commentary. In the simulations that follow, the model sets an initial operating point that balances the pressures from safety and efficiency so that the model starts in equilibrium conditions.

The model allows us to isolate the effects of the two causal mechanisms—migration from efficiency pressure and excursions from experimentation. To examine the effects of excursions, Fig. 8 shows the results of simulations in which the efficiency goal is held constant. The economic failure boundary is held fixed, and the efficiency goal is a constantly desired distance from the economic boundary.

Fig. 8 Baseline behavior when efficiency goal is constant

The lower panel of Fig. 1 depicts Cook and Rasmussen's representation of three safety systems that operate in different regimes: the stable low risk system, the stable high risk system, and the unstable system. The simulations in Fig. 8 faithfully replicate the behavior of the three types of systems described in Cook and Rasmussen (2005). A stable low-risk system, as in the upper left quadrant, operates far from the acceptable boundary. Even with larger excursions, as in the upper right quadrant, a stable low-risk system maintains its distance from the acceptable boundary. A stable high-risk system, such as a high-reliability organization (HRO), as shown in the lower left quadrant, operates closer to the acceptable boundary but takes smaller excursions and stays within the safe operating region (Weick et al. 1999). An unstable high-risk system, shown in the lower right quadrant, has much larger and more rapid shifts in the operating point.

The simulations above establish the baseline behavior of the model and constitute a behavior replication test (Sterman 2000). By replicating the patterns of behavior described by Cook and Rasmussen (2005), they establish the usefulness

of the model. We now turn to additional simulations to explore the behavior of the dynamic safety system, building on and extending Rasmussen's work. In our next set of experiments, we allow the economic failure boundary to migrate as the system adjusts to the experience of operating performance over time. When the operating point achieves greater efficiency (lower values in our diagrams) without an instance of safety breach, the organizational norm adjusts to expect greater future levels of efficiency. Market forces also push the boundary in the direction of greater efficiency, as customers demand more and competitors provide alternatives that meet these demands. Figure 9 shows simulations in which the Economic Failure Boundary varies endogenously for four different values of the standard deviation of the excursions.

Figure 9 also includes some analytical metrics used to describe the simulation runs. First fail is the first time at which the operating point actually strays into the unsafe region, denoting an accident or safety breach. This will occur when an excursion takes the system not only past the marginal boundary but also past the acceptable performance

Fig. 9 Safety dynamics with adaptive economic failure boundary for various standard deviations

boundary (here set to the value 0). Total fails is the number of times that the operating point actual is beyond the acceptable performance boundary and is thus a measure of the number of accidents or safety breaches over the time horizon of the simulation (= 100 weeks in these simulations). Total failure is based on sampling in every time step. The model also captures another measure of accident frequency called fail instances which excludes any excursion beyond the acceptable performance boundary if the operating point actual in the previous time step was also beyond the boundary.

The simulations in Fig. 9 allow the economic failure boundary to vary endogenously according to the recent performance of the organization. The effect is symmetric in the sense that increases and decreases in the economic failure boundary are equally easy to effect. A more likely case is that management and market forces will push towards greater efficiency when experience shows that it is possible and will be reluctant to relax the boundary based on a short period of poor performance. Indeed, there is little

evidence that market forces would accommodate such a relaxation. Rasmussen is somewhat silent about the migration of the economic boundary, although he does note that “defences are likely to degenerate systematically through time when *pressure towards cost-effectiveness is dominating*” and that “the normal efforts of many actors in their respective daily work contexts, responding to the *standing request to be cost-effective*” can set the stage (Rasmussen 1997: p 190) (*italics added*). Dominating pressure and the standing request for cost-effectiveness are consistent with a conceptualization that adapts the economic failure boundary asymmetrically, favoring cost-effectiveness. To incorporate this additional realism in our model, we modify the formulation for determining the economic failure boundary so the adjustment (modeled by the variable change in economic boundary) updates only in the direction of greater efficiency. This formulation resembles remembering how efficient the system has been in the past and holding onto the expectation to repeat such historical efficiencies. Figure 10 repeats the

Fig. 10 Safety dynamics with asymmetric adaptive economic failure boundary for various standard deviations

simulations of Fig. 9 using this more realistic representation of the efficiency gradient.

The simulations in Fig. 10 highlight several important features of the dynamic safety environment. First, as we see in the upper left panel of Fig. 10, the system can operate in a safe regime for extended periods of time. When excursions from the normal operating point are small in magnitude, there is a much smaller likelihood of crossing the acceptable performance boundary. While this may seem obvious at first glance, a more nuanced look reveals that this safety record is achieved in the system because there is a balance between the pressures for efficiency and the pressures for safety that keep the operating point at some distance from the acceptable performance boundary. Second, the large excursions have two consequences. A direct effect is that they increase the risk of accidents. Large excursions, which occur in the simulations as the standard deviation is increased, lead to a higher risk of crossing the acceptable performance boundary. A second effect is that larger excursions lead to faster and greater migration of the economic failure boundary leading to more efficiency pressure. Excursions that do not end in safety breaches “teach” the system (both the internal workforce and the external customers and industry) that such excursions are possible, and performance expectations adjust. Third, the safety pressure (counter gradient from campaigns for safety) exerts its influence on the normal operating point (the kernel) but it has little influence on the excursions. The nature of the excursions is that they arise from experiments or mishaps that introduce variations even in the face of the pressure to maintain safety.

The lower right panel of Fig. 10 can be interpreted as an example of the phenomenon of “going solid”. Going solid is a slang term from the nuclear power industry that refers to a technical situation that is difficult to manage, such as when the contents of a boiler turn from a combination of water and steam to all water (US_Nuclear_Regulatory_Commission 2004). When a system transitions from loosely coupled

to tightly coupled, “the increased complexity ... makes 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,” a situation Cook and Rasmussen (2005, p.132) refer to as “going solid”. As we see in the lower right panel, such a system is more prone to failure.

We examine one possibly stabilizing policy, that of more tightly holding on to the Marginal Boundary to prevent drift. Rather than allow this boundary to loosely drift as a result of excursions, we examine a policy that introduces a perception delay to the updating of this marginal boundary. We operationalize this policy in the model by changing the input for the change in marginal boundary flow from the operating point actual to the perceived operating point. In Table 3 below, we present the results of this policy as the “Sticky” safety boundary. Although this policy has a significantly stabilizing effect, it does not eliminate the safety challenges that we see in the previous simulations.

The simulations displayed in the previous figures each show the results from only one simulation run for a particular realization of the random stream modeling variability in excursions. We conducted 1000 simulations of each of the scenarios from Figs. 8, 9 and 10 and tabulated the results in Table 3.

8 Discussion

Rasmussen’s conceptual model is a verbal description of forces (i.e., pressures or gradients) that generate a drift in the boundary of acceptable performance. His paper and subsequent commentaries describe this drift of the boundary and other behavior of the conceptualized system. Thus, we have a verbal description of a system structure coupled with a verbal description of system behavior the structure might generate. Together, this structure and behavior constitute

Table 3 Statistics from 1000 simulations for each scenario

		Std Dev = 1			Std Dev = 3			Std Dev = 4			Std Dev = 5		
EBF Adjustment		Sym	Asym	Asym	Sym	Asym	Asym	Sym	Asym	Asym	Sym	Asym	Asym
“Sticky” safety boundary		No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
First fail	Mean	0.00	0.00	0.00	28.98	43.53	37.82	23.99	16.58	19.44	10.67	8.33	9.15
	Std Dev	0.00	0.00	0.00	33.88	26.11	31.51	19.41	12.06	15.66	9.09	5.98	7.37
Failure count	Mean	0.00	0.00	0.00	1.21	6.66	3.29	9.48	29.00	21.61	23.72	61.59	53.56
	Std Dev	0.00	0.00	0.00	1.64	4.99	3.59	7.45	16.46	31.14	20.98	31.50	52.09
	% Reduction	0.00	0.00	0.00			51%			25%			13%
Unique failures	Mean	0.00	0.00	0.00	0.77	3.32	1.75	4.76	11.81	8.75	10.24	21.50	18.33
	Std Dev	0.00	0.00	0.00	0.87	2.09	1.62	2.25	4.01	3.59	3.98	5.46	5.42
	% Reduction	0.00	0.00	0.00			47%			26%			15%

a dynamic hypothesis (Randers 1980; Sterman 2000). In this paper, we have translated his theory into a formal system dynamics model—a mathematical formulation of the conceptual model that Rasmussen presented in his 1997 paper—and used the model to test the dynamic hypothesis and to conduct experiments to explore the behavior of the system. These results highlight several important features of the dynamic safety problem.

First, the most basic question we asked is whether Rasmussen's theory can indeed account for the behavior of organizational drift that it claims to explain. His theory describes a feedback process that we have first diagrammed, then formalized mathematically, and then tested using simulation of the dynamics. We used the causal loop diagram to understand the key feedback process. As the operating point moves due to natural variations in activity, excursions may result in a decrease (movement in the less safe direction) in the operating point actual which in turn causes a decrease in the marginal boundary. A decrease in the marginal boundary causes a decrease in the safety goal, resulting in a decrease in safety gradient pressure. With less pressure from the safety gradient, the less constrained operating point kernel also decreases. Excursions from this new, lower operating point kernel generate an operating point actual that is still lower, and the cycle continues. The original downward change in the operating point kernel is continued or reinforced by the loop, so it is a reinforcing feedback loop. Moving from the causal loop diagram to the formal model, the simulation analysis demonstrates that the structure can indeed generate the patterns of behavior he posited. As a test of the dynamic hypothesis, these simulations give support to Rasmussen's theory.

Second, the explicit representation of the feedback process also calls attention to the two effects of an excursion towards the boundary of acceptable performance. First, the lower operating point actually leads to updating both the operating point kernel, (essentially the memory of recent operating point experience) and the marginal boundary towards the boundary of acceptable performance. The second effect is that as the operating point kernel moves towards the boundary of acceptable performance, it also decreases the safety gradient pressure, so there is less resistance to the drift. The result is a vicious cycle that if left unchecked would propel the organization past the marginal boundary. Rasmussen's original description did not explicitly identify the second effect, although the reinforcing nature of the feedback process is rooted in this latter effect. One significant feature of reinforcing loops is their tendency to be self-sustaining. However, they also need a form of stimulus to get them moving, so one important strategy for managing such situations is to put in place measures to prevent such stimuli from "jump starting" the loop. The focus of high-reliability organizations on tight control of the operating point, even

when operating near the margin, is an example of such a prevention strategy (Cook and Rasmussen 2005).

Third, the process of formalizing the theory revealed a need for further specificity to define the pressures resulting from the gradients for safety and for efficiency. Rasmussen's original theory provided an adequate description of the *directions* of these gradients but not so for the *magnitudes* of the gradients. It is quite common when formalizing a theory to discover gaps in the specification of the theory (Peli et al. 1994). In this case, the issue was a missing construct. To fully specify the operational nature of the gradients, we included a new construct that we labeled the safety goal (or efficiency goal). While there could be a special case in which the safety goal (Efficiency goal) was identical to the marginal boundary (Economic failure boundary), in the more general case, the implicit target for the organization is to operate within the boundary.

There are practical implications of including the safety goal in our formalization. First, the existence of the safety goal as another real-world component of the mechanism that yields organizational drift implies there may be additional possible management tactics based on it. Second, while the safety goal and the boundaries are useful theoretically, characterizing the real world equivalents will generally require multi-dimensional definitions, and the dimensions would likely span a range of both technical and social domains. Compromising the safety goal along any dimension, or some combination of dimensions, could lead to a breach of the boundary of acceptable performance and an accident. Third, an important difference between the safety goal and the marginal boundary relates to knowability. Although there can be cases in which the safety goal is implicitly defined by the safety gradient pressure, more generally the safety goal can be made explicit or more precisely explicit along at least some dimensions. As Cook and Rasmussen explain, the location of the boundary of acceptable performance is generally not knowable, because the only way to truly discover it is to pass it and suffer an accident. And, even in the aftermath of an accident, it will not likely be clear by how much the boundary was violated. In contrast, the safety goal can be known because it will be specified and consequently it can serve as a means of monitoring drift. Although the organization might not know the margin between its safety goal and the boundary of acceptable performance, a comparison of the current safety goal with the safety goal of an earlier time period can reveal migration and trigger possible intervention to reverse the trajectory towards greater risk.

A fourth feature of the dynamic safety problem revealed through the simulation analysis goes beyond the original work of Rasmussen and Cook. Whereas Rasmussen's theory focused on the drift of the safety boundary, our simulations introduced scenarios that generated drift in the economic failure boundary. Processes analogous to the

ones that cause drift in the marginal boundary in the direction towards less safety can cause drift in the economic failure boundary, also in the direction towards less safety. By moving from Rasmussen's conceptualization of a fixed economic failure boundary to one that allows for dynamic updating of the boundary, we identified an additional, perhaps more dangerous form of organizational pathology. Our simulations demonstrate the problematic nature of the drifting economic failure boundary that results from endogenous forces: managerial and investor expectations for economic performance are heavily conditioned by demonstrated efficient performance. Strong economic performance, even if achieved with strategies with high risk for safety breaches, sets the target for similar or stronger economic performance in the future. Moreover, in addition to the endogenous processes in our model, there are exogenous forces that conspire to redefine the location of the economic failure boundary such as changing market conditions, innovations in technology and product design, pressure from the improving performance of competitors, and more demanding consumer expectations. The drifting safety boundary reflects the organization's perception of the feasibility of the riskier operating point, while the drifting economic boundary reflects the perception of the desirability of the more profitable operating point. The potential negative consequences of the combination of these sources of drift suggest that under conditions such as the exogenous forces listed previously, managers and workeres should be especially diligent to guard against safety breaches or high variability in excursions.

As with all models, the model developed here is a simplification that intentionally omits much. For example, the current model does not separate multiple gradients of pressure to drift towards less safety nor does it describe the boundaries and safety goal in more than one dimension. If these were added to the model, we might see a closer correspondence with practitioner language for descriptions of risk, safety, and other factors but it is not likely that we would see meaningful changes in the dynamic behavior. Instead, we have seen that Rasmussen's original, concise theory provided fertile ground for exploring the dynamic safety problem and specifically the phenomenon of organizational drift, enabling experiments to extend the theory (e.g., by allowing the economic failure boundary to drift as well), generating new insights, and discovering practical implications for managing such systems. Moreover, the model opens the door for further experimentation that has the potential to lead to refinements and elaboration of the original theory. System dynamics contributed a useful methodology by highlighting the role of the feedback structure in generating patterns of dynamic behavior, enabling the formalization of the theory, and providing the means for simulation analysis. The present study suggests

that formalizing and simulating the implications of existing theory can help to provide more nuanced understanding, greater confidence in the theory's ability to explain behavior, a broader scope of applicable scenarios or conditions, and a richer set of priorities for managing a dynamic safety problem.

Acknowledgements The authors gratefully acknowledge the support provided by the National Institutes of Health Grant 5R01GM105049-02 from the National Institute of General Medical Sciences.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Axelrod R (2007) Simulation in social sciences. Handbook of research on nature-inspired computing for economics and management. IGI Global, Hershey, pp 90–100
- Ben-Daya M, Raouf A (1996) A revised failure mode and effects analysis model. *Int J Qual Reliab Manag* 13(1):43–47
- Black LJ, Carlile PR, Repenning NP (2004) A dynamic theory of expertise and occupational boundaries in new technology implementation: building on Barley's study of CT scanning. *Adm Sci Q* 49(4):572–607
- Bouloiz H, Garbolino E, Tkiouat M, Guarnieri F (2013) A system dynamics model for behavioral analysis of safety conditions in a chemical storage unit. *Saf Sci* 58:32–40
- Carley KM, Wallace WA (2001) Computational organization theory. Springer, Berlin
- Cook R, Rasmussen J (2005) "Going solid": a model of system dynamics and consequences for patient safety. *Qual Saf Health Care* 14(2):130–134
- Cooke DL (2003) A system dynamics analysis of the Westray mine disaster. *Syst Dyn Rev J Syst Dyn Soc* 19(2):139–166
- Cooke DL, Rohleder TR (2006) Learning from incidents: from normal accidents to high reliability. *Syst Dyn Rev* 22(3):213–239
- Cyert R, March J (1963) A behavioral theory of the firm. Prentice Hall, Englewood Cliffs
- Davis JP, Eisenhardt KM, Bingham CB (2007) Developing theory through simulation methods. *Acad Manag Rev* 32(2):480–499
- Dekker S (2016) Drift into failure: from hunting broken components to understanding complex systems. CRC Press, Boca Raton
- Dekker SWA (2011) Drift into failure: from hunting broken components to understanding complex systems. Ashgate, Farnham
- Farjoun M (2005) Organizational learning and action in the midst of safety drift: Revisiting the space shuttle program's recent history. In: Starbuck WH, Farjoun M (eds) *Organization at the Limit: Lessons from the Columbia Disaster*. Blackwell, Malden, MA, pp 60–80
- Forrester JW (1961) Industrial dynamics. Productivity Press, Cambridge

- Galuska PA (2012) *Thunder on the mountain: death at massey and the dirty secrets behind big coal*. St. Martin's Press, New York
- Goh YM, Love PE (2012) Methodological application of system dynamics for evaluating traffic safety policy. *Saf Sci* 50(7):1594–1605
- Hall RI (1976) A system pathology of an organization: the rise and fall of the old saturday evening post. *Adm Sci Q* 21(2):185–211
- Hannan MT, Freeman J (1984) Structural inertia and organizational change. *Am Sociol Rev* 49:149–164
- Harrison JR, Lin Z, Carroll GR, Carley KM (2007) Simulation modeling in organizational and management research. *Acad Manag Rev* 32(4):1229–1245
- Heinrich HW (1931) *Industrial accident prevention: a scientific approach*. McGraw-Hill, New York
- Hollnagel E (1983) Position paper for NATO conference on human error. In: Senders JW, Moray N (eds) *NATO conference on human error*, vol 7. OECD Halden Reactor Project, Bellagio
- Hollnagel E (1991) Does human error exist? In: Senders JW, Moray NP (eds) *Human error: cause, prediction, and reduction*, vol 153. Lawrence Erlbaum Associates, Hillsdale
- Hopkins A (2008) *Failure to learn: the BP Texas city refinery disaster*. CCH, Australia
- Hopkins A (2018) *Disastrous decisions: the human and organisational causes of the Gulf of Mexico blowout*. CCH, Australia
- Hurwitz B, Sheikh A (2009) Medical fallibility—cultural recognition and representation. *J R Soc Med* 102(5):181–185
- Klein G, Orasanu J, Calderwood R, Zsombok CE (eds) (1993) *Decision making in action: models and methods*. Ablex Publishing Company, Norwood
- Levinthal D (1997) Adaptation on rugged landscapes. *Manag Sci* 43(7):934–950
- Liu Q, Li X, Hassall M (2015) Evolutionary game analysis and stability control scenarios of coal mine safety inspection system in China based on system dynamics. *Saf Sci* 80:13–22
- Lopes LL (1986) Aesthetics and the decision sciences. *Syst Man Cybern IEEE Trans* 16(3):434–438
- March JG (1991) Exploration and exploitation in organizational learning. *Organ Sci* 2(1):71–87
- Martinez-Moyano IJ, McCaffrey DP, Oliva R (2014) Drift and adjustment in organizational rule compliance: explaining the “regulatory pendulum” in financial markets. *Organ Sci* 25(2):321–338
- Masuch M (1985) Vicious circles in organizations. *Adm Sci Q* 30:14–33
- Merton R (1948) The self-fulfilling prophecy. *Antioch Rev* 8(193–210):193–210
- Minami NA, Madnick S (2009) Dynamic analysis of combat vehicle accidents. *Syst Dyn Rev J Syst Dyn Soc* 25(2):79–100
- Mitropoulos P, Abdelhamid TS, Howell GA (2005) Systems model of construction accident causation. *J Constr Eng Manag* 131(7):816–825
- Mohamed S, Chinda T (2011) System dynamics modelling of construction safety culture. *Eng Constr Archit Manag* 18(3):266–281
- Moray N (1988) *Ex Risø semper aliquid antiquum: sources of a new paradigm for engineering psychology*. In: Goodstein LP, Andersen HB, Olsen SE (eds) *Tasks, errors, and mental models: a festschrift to celebrate the 60th birthday of Professor Jens Rasmussen*. Taylor & Francis, London, pp 12–17
- Morecroft JDW (1985) Rationality in the analysis of behavioral simulation models. *Manag Sci* 31(7):900–916
- Morrison JB (2012) Process improvement dynamics under constrained resources: managing the work harder versus work smarter balance. *Syst Dyn Rev (Wiley)* 28(4):329–350
- Morrison JB, Oliva R (2018) Integration of behavioral and operational elements through system dynamics. *Handb Behav Oper* 8287
- Nelson R, Winter SG (1974) Neo-classical vs. evolutionary theories of economic growth. *Econ J* 84:886–905
- Nelson RR, Winter SG (1982) *An evolutionary theory of economic change*. Harvard University Press, Cambridge
- Peli G, Bruggeman J, Masuch M, Nuallain BO (1994) A logical approach to formalizing organizational ecology. *Am Sociol Rev* 59(4):571–593
- Perrow C (1984) *Normal accidents: living with high risk systems*. Basic Books, New York
- Rahmandad H (2008) Effect of delays on complexity of organizational learning. *Manag Sci* 54(7):1297–1312
- Randers J (1980) Guidelines for model conceptualization. In: Randers J (ed) *Elements of the system dynamics method*. Productivity Press, Cambridge, pp 117–139
- Rasmussen J (1983) Skills, rules and knowledge; signals, signs and symbols and other distinctions in human performance models. *IEEE Trans Syst Man Cybern* 13(3):257–266
- Rasmussen J (1985) The role of hierarchical knowledge representation in decision making and system management. *IEEE Trans Syst Man Cybern Part A Syst Hum* 15:234–243
- Rasmussen J (1990) The role of error in organizing behaviour. *Ergonomics* 33:1185–1190
- Rasmussen J (1993) Diagnostic reasoning in action. *IEEE Trans Syst Man Cybern Part A* 23(4):981–992
- Rasmussen J (1997) Risk management in a dynamic society: a modelling problem. *Saf Sci* 27(2):183–213
- Rasmussen J, Batstone R (1989) Why do complex organizational systems fail? 45. The World Bank, Washington, DC
- Rasmussen J, Duncan K, Leplat J (eds) (1987) *New technology and human error*. Wiley, Chichester
- Rasmussen J, Nixon P, Warner F (1990) Human error and the problem of causality in analysis of accidents [and discussion]. *Philos Trans R Soc Lond Ser B Biol Sci* 327(1241):449–462
- Reason J (1990) *Human error*. Cambridge University Press, Cambridge
- Reason J (1997) *Managing the risks of organizational accidents*. Ashgate Publishing Limited, Alersshot
- Richardson GP (1991) *Feedback thought in social science*. University of Pennsylvania Press, Philadelphia
- Richardson GP (2011) Reflections on the foundations of system dynamics. *Syst Dyn Rev* 27(3):219–243
- Richmond B (1993) Systems thinking: critical thinking skills for the 1990s and beyond. *Syst Dyn Rev (Wiley)* 9(2):113–133
- Rudolph JW, Morrison JB, Carroll JS (2009) The dynamics of action-oriented problem solving: linking interpretation and choice. *Acad Manag Rev* 34(4):733–756
- Rudolph JW, Repenning NP (2002) Disaster dynamics: understanding the role of quantity in organizational collapse. *Adm Sci Q* 47(1):1–30
- Sastry MA (1997) Problems and paradoxes in a model of punctuated organizational change. *Adm Sci Q* 42:237–275
- Shin M, Lee H-S, Park M, Moon M, Han S (2014) A system dynamics approach for modeling construction workers' safety attitudes and behaviors. *Accid Anal Prev* 68:95–105
- Snook SA (2000) *Friendly fire: the accidental shoot-down of US black hawks over Northern Iraq*. Princeton University Press, Princeton
- Starbuck WH, Farjoun M (2005) *Organization at the limit: lessons from the columbia disaster*. Blackwell, Malden
- Sterman J, Oliva R, Linderman K, Bendoly E (2015) System dynamics perspectives and modeling opportunities for research in operations management. *J Oper Manag* 39–40:1–5
- Sterman JD (1994) Learning in and about complex systems. *Syst Dyn Rev* 10(2):291–330
- Sterman JD (2000) *Business dynamics: systems thinking and modeling for a complex world*. Irwin-McGraw Hill, Chicago
- Sterman JD, Repenning NP, Kofman F (1997) Unanticipated side effects of successful quality programs: exploring a paradox of organizational. *Manag Sci* 43(4):503
- Tushman ML, Romanelli E (1985) Organizational evolution: a metamorphosis model of convergence and reorientation. In: Cummings

- BMS (ed) Research in organizational behavior, vol 7. JAI Press, Greenwich, pp 171–222
- Tversky A, Kahneman D (1974) Judgement under uncertainty: heuristics and biases. *Science* 185(4157):1124–1131
- US_Nuclear_Regulatory_Commission (2004) 25th Anniversary of three mile island unit 2 presentation. Rockville, MD
- Vaughan D (1996) The challenger launch decision: risky technology, culture and deviance at NASA. University of Chicago Press, Chicago
- Wears RL (2017) Rasmussen number greater than one. *Appl Ergon* 59(Part B):592–597
- Weick KE (1987) Organizational culture as a source of high reliability. *Calif Manag Rev* 29(2):112–127
- Weick KE, Sutcliffe KM, Obstfeld D (1999) Organizing for high reliability: processes of collective mindfulness. *Res Organ Behav* 21:81–123
- Woods DD (2005) Creating foresight: lessons for enhancing resilience from Columbia. In: Starbuck WH, Farjoun M (eds) *Organization at the limit: lessons from the Columbia disaster*. Blackwell, Malden
- Woods DD, Dekker SWA, Cook RI, Johannesen L, Sarter N (2010) *Behind human error*, 2nd edn. Ashgate, Farnham
- Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. *J Comp Neurol Psychol* 18(5):459–482

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.