Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Reviews of resilience theories and mathematical generalization

2.1 Resilience theories and practices in socioecological systems

Abbreviations

| Abbreviation | Description                  |
|--------------|------------------------------|
| SES          | Socioecological system       |
| CAS          | Complex adaptive system      |
| CFCs         | Chlorofluorocarbons          |
| SARS         | Severe acute respiratory syndrome |
| NGO          | Nongovernmental organization |

In this section, the resilience approach in socioecological systems (SESs) is introduced first, including its underlying rationales and assumptions. Then, the key principles for enhancing the capacity of SES to deal with change and disturbance are discussed, along with their social and political implications. For more detailed discussions, refer to the original publication (Biggs et al., 2015a).

2.1.1 The resilience approach

SES research is a fast-growing interdisciplinary field of study, including the social, economic, political and ecological sciences. The resilience approach focuses specifically on the capacity of SES to deal with change, particularly unexpected change in these kinds of systems. Here, resilience of an SES is defined as the capacity of an SES to sustain human well-being in the face of change, both by buffering shocks but also through adapting or transforming in response to change.

Since the beginning of the 20th century, mankind has experienced unparalleled rapid development due to huge advances in science and technology, manifested by a 15-fold increase in the global economy, large-scale conversion of land to agriculture and an increase in the global population from 1.6 billion in 1900 to over 7 billion in 2011 (MA, 2005; Steffen et al., 2007). The rapid changes have brought huge benefits and dramatic improvements to people’s lives. Consider the average life expectancy for example. For most of human history the figure stood at 20–30 years, reflecting the combined effects of poor nutrition, disease, warfare, and especially the low infant survival rate (Lancaster, 1990). In 2010, the average life
expectancy had reached 67 years, which was an astonishing feat when considering the fact that in 1900 it was merely 31 years (CIA, 2013). And it is predicted to continue increasing.

Although the command of modern technology has enabled mankind to cope better with nature, it does not put human beings above nature. Ultimately people still depend largely on nature for a variety of essential needs, including fresh air, clean water and food, protection from hazards such as droughts and storms, and a wide variety of cultural, spiritual and recreational needs that play a key role in human well-being. Such benefits derived from the interaction of people with nature are known as ecosystem services (Ernstson, 2013; Reyers et al., 2013; Huntsinger and Oviedo, 2014). Ecosystem services can be divided into three categories: provisioning services such as food and fuel, regulating services such as disease and flood control, and cultural services such as sports and education.

It has become increasingly evident that the massive scale and extent of human activities such as agriculture, transport, release of novel chemicals, and various environmental problems are degrading the capacity of nature to generate key ecosystem services on which we depend. Climate change provides a prime example. It is scientifically proven that rising atmospheric carbon dioxide levels resulting from anthropogenic fossil-fuel combustion and land clearing are changing rainfall and temperature patterns around the globe and leading to an increased incidence of extreme events such as droughts and storms (IPCC, 2014). Other undermining effects on nature include biodiversity loss and changes in nutrient cycles. All these changes are impacting food security, disease prevalence and infrastructure, as well as affecting traditional lifestyles and cultural practices that shape people’s identity.

Human-induced changes to the environment are also increasing the risk of crossing critical thresholds or tipping points in the ecosystem that could lead to large, nonlinear and potentially irreversible changes at local through global scales, such as the death of coral reefs, shifts in regional monsoon rainfall patterns or the collapse of the Greenland ice sheet (MA, 2005; Rockström et al., 2009; Barnosky et al., 2012). Beyond these somewhat known effects, our impacts on the environment are also leading to completely novel changes that are very difficult to anticipate, and could have dramatic impacts on a variety of ecosystem services. The formation of the ozone hole due to the use of chlorofluorocarbons in refrigerators (Farman et al., 1985), the potential emergence and spread of new diseases such as severe acute respiratory syndrome, and the potential consequences of nuclear proliferation and massively increased global connectivity and trade on the environment are all grave concerns (Martin, 2007).

In the face of these rapid, ongoing changes to the environment and human society, continued human well-being obviously requires breakthrough approaches to cope with change, especially unexpected change, as part of their normal functioning. In the words of Holling, the founder of resilience theory in ecology, It is that creation of something fundamentally novel that gives an evolutionary character to development of a region that might make sustainable development an achievable reality rather than an oxymoron (Holling, 1996). As Holling showed us in his epochal review paper published in 1973, “Resilience and stability of ecological
systems” (Holling, 1973), nature’s reaction to change teaches us the best lessons. Based on his shrewd observation, he offered two different interpretations of the behavior of ecological systems:

*The whole sequence of environmental changes can be viewed as changes in parameters or driving variables, and the long persistence in the face of these major changes suggests that natural systems have a high capacity to absorb change without dramatically altering. . . . . It is useful to distinguish two kinds of behaviour. One can be termed stability, which represents the ability of a system to return to an equilibrium state after a temporary disturbance; the more rapidly it returns and the less it fluctuates, the more stable it would be. But there is another property, termed resilience, that is a measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables.*

The resilient property of nature best exemplifies system resilience to change and disturbance. Since the publication of Holling’s resilience theory, the resilience approach has increasingly attracted researchers’ attention, beginning with ecological studies and later expanding to socioecological studies, computer science, engineering and civil infrastructural systems, focusing on system resilience against change, expected or unexpected, incremental or abrupt. In general, resilience research is about studying system attributes that may promote or undermine the resilience of a system, identifying key principles for building resilience and establishing practical procedures for their applications in real-world settings.

Fundamental to the resilience approach of SES is the view that human beings are integrated into the biosphere, or the global ecological system at local to global scales and their activities help shape their environment, which they are fundamentally dependent on for various ecosystem services that underpin human well-being (Berkes and Folke, 1998; Berkes et al., 2003; Walker and Salt, 2006). These intertwined SESs are assumed to be complex adaptive systems (CASs), i.e., they have the capacity to self-organize and adapt based on past experience and changing conditions, and are characterized by nonlinear dynamics that may generate uncertainties regarding system behavior (Norberg and Cumming, 2008). Consequently, in SES the resilience approach to change involves not only recovery from unexpected shocks and avoiding undesirable thresholds or tipping points, but also the capacity to adapt to ongoing change and fundamentally transform SES if necessary (Walker et al., 2004; Folke et al., 2010). The resilience approach is considered to lie within the broad emerging field of sustainability science, which studies the interactions between nature and society for long-lasting human well-being without compromising the social, economic and environmental foundations.

Although resilience theories in SESs have not yet reached the stage of maturity, a certain common sense has been built on several theorems or principles that are considered important to building system resilience in SES. The following seven principles have been selected by a group of young resilience scholars from a wide range of disciplinary backgrounds, forming a tentative theoretical framework for developing the resilience theories in SES. These are: Principle 1—Maintain
diversity and redundancy; Principle 2—Manage connectivity; Principle 3—Manage slow variables and feedbacks; Principle 4—Foster CAS thinking; Principle 5—Encourage learning; Principle 6—Broaden participation; Principle 7—Promote polycentric governance; refer to Fig. 2.1. The first three principles are about generic SES properties and processes that enhance resilience, and the last four focus on the way in which SESs are governed. For more detailed discussions, refer to Biggs et al. (2015b).

2.1.2 Principle 1—maintain diversity and redundancy

**Definitions of diversity and redundancy**

According to Stirling (2007), diversity has three interrelated but distinct aspects common to diversity concepts across many different fields, i.e., variety (how many different types of element), balance (ratio of each element), and disparity (degree of differences between elements). Depending on the elements being considered, different types of diversity may exist within a particular SES. For instance, functional diversity among plant or animal species may be described by the number of species present (variety), their relative abundance (balance) and the extent to which the species differ from each other (disparity) (Kotschy, 2013). Likewise, cultural diversity may be described by the number of cultural groups present (variety), their relative size or influence (balance), and the differences between them (disparity). As resilience is about coping with change, the differences in response of SES elements to
disturbance (response diversity) are considered to be particularly important for resilience (Elmqvist et al., 2003; Leslie and McCabe, 2013; Mori et al., 2013).

Redundancy represents the replication of elements or pathways in a system and is determined by the number of elements that perform a particular function similarly (Walker, 1992). Redundancy provides potential means for replacing failed or lost elements, thus increasing system reliability. In general, elements that show different responses to disturbances provide response diversity, while elements that perform a particular function similarly provide redundancy for that function.

**Resilience-enhancing mechanisms**

Ecosystem services are viewed as coproduced by the ecological and social components of an SES (Reyers et al., 2013). Take fish or vegetables that we eat as an example of a provisioning ecosystem service, which are produced by a combination of ecological processes and the human knowledge and skills to catch or cultivate, store and transport the produce to the place where it is consumed. The resilience of an ecosystem service may be influenced by redundancy and diversity pertaining to the system elements involved in producing that ecosystem service.

Redundancy due to the presence of multiple elements with similar functional roles allows for substitution among elements when the failure or loss of elements occurs, thus ensuring functional continuity in an SES. In the case of farming, to ensure a continuous food supply it is common practice for poor small-scale farmers to plant several different types of food crop to reduce the potential impact on food provision caused by failure of any one crop (Altieri, 2009). The fact that several crops contribute to the provision of food and can be substituted for each other provides redundancy in food supply for these farmers. It is known that different crop types or species often respond differently to disturbances such as drought or disease (response diversity) and hence simultaneous failure of all crops is unlikely. Therefore, in this case the redundancy and diversity in food supply for these farmers enhance the resilience of this particular provisioning ecosystem service.

Take the social governance system as another example. The existence of a variety of organizational forms such as government department, nongovernmental organization and community organization with overlapping domains of authority can provide both redundancy and response diversity in governing or regulating the usage of ecosystem services. In regulating the use of a particular ecosystem service, if different organizations all play a similar role, redundancy is present because the regulatory function will continue even if one or more of the organizations become dysfunctional. As organizations with different sizes, cultures, funding mechanisms and internal structures are likely to respond in different ways to various socioecological challenges, these differences between organizational forms and structures naturally lead to response diversity (Williamson, 1985; Ostrom, 2005). Hence, if managed properly, the functional redundancy and response diversity in the social
governance system are likely to enhance the system’s adaptive capacity and resilience for maintaining ecosystem services and dealing with disturbance and change.

**Related issues**

Several issues concerning diversity and redundancy should be addressed. It is known that in some SESs, high redundancy and low response diversity may coexist. This happens when there are many system elements contributing to a particular ecosystem service, but all the elements have similar structures either by design (e.g., human institutions or activities) or due to environmental or historical constraints. This kind of system may function well when change and disturbance occur within the traditional range of variation experienced by the system, and thus resilience is expected for the provision of ecosystem services. However, when facing new types of disturbance, such as those caused by climate change, the system is likely to be vulnerable because of the lack of options available for responding to the new challenges (Janssen et al., 2007).

Also, a very high level of diversity and redundancy can sometimes have negative impacts on the resilience of ecosystem services, because it increases the possibility for inactivity within some aspects of SES. In management organizations, high redundancy may hinder governance functions for ecosystem services, because it tends to increase not only administrative costs, but also the potential for power struggles and contradictory regulations that are known to compromise the ability of governance systems to respond effectively to change (Jentoft et al., 2009). These problems should be noted when studying diversity and redundancy for enhancing the resilience of ecosystem services in SES. In ecological systems, there is less evidence of negative influences of high diversity and redundancy on system resilience to change and disturbance. For more detailed discussions, refer to Kotschy et al. (2015).

### 2.1.3 Principle 2—manage connectivity

**Definition of connectivity**

Connectivity is defined as the way in which parts of an SES (e.g., species, landscape patches) interact with each other (e.g., exchanging information, transferring material). If a system can be considered as being composed of individual components, then connectivity refers to the nature and the strength of the interactions between these components. From a network perspective, all individual components of a system are nodes integrated into a web of connections that form the links. Examples of links are species interactions (e.g., the mutualism interaction between flowering plants and their animal pollinators), vegetation corridors across habitats or communication channels between human communities. The structure of an SES is determined by how the links are arranged within the system, which includes not only the structure of connections between components (e.g., with or without links, one-way, or mutual links), but also the strength of these links.
**Resilience-enhancing mechanisms**

In general, connectivity in SES provides the pathways for the flow of energy, material, or information necessary for building the resilience of ecosystem services. Depending on its structure and strength, connectivity may facilitate recovery or constrain locally the spread of a disturbance, hence safeguarding ecosystem services against change (Nyström and Folke, 2001).

The recolonization of coral reefs is a good example that demonstrates the importance of connectivity for the recovery of disturbed SESs and thus for the maintenance of their ecosystem-service resilience. Studies show that the degree of connectivity between remnant coral patches influences the extent of reef recolonization, and this connectivity is determined by the prevailing currents that allow coral recruitment between neighboring reefs (Treml et al., 2007; Mumby and Hastings, 2008). Similarly, in disturbance experiments, scientists found that recovery of macrobenthic communities was largely determined by the degree of connectivity across metacommunities (Thrush et al., 2008). Well-connected habitats with no physical barriers facilitate the recolonization of nearby sites, making loss of species due to change and disturbance less likely than poorly-connected habitats. This is why in many conservation initiatives that aim at enhancing the resilience of SES, approaches that protect, maintain, and restore connectivity between habitats are given priority.

In some cases, connectivity may enhance the resilience of ecosystem services not by facilitating recovery from a disturbance, but by acting as a barrier to the spread of disturbances (e.g., the spread of fire or of disease vectors). For instance, when local losses of ecosystem services due to local disturbance become inevitable, limited connectivity with the outside world can reduce the possibility of large-scale global effects, i.e., containing the potential loss of ecosystem services to a local scale. This function of connectivity is frequently utilized in fighting the spread of disease, wildfire, or nuclear contamination by isolating the affected area.

In general, connectivity facilitates the maintenance of biodiversity in the landscape, which is vital to the production of many ecosystem services, hence indirectly enhancing the resilience of SES. For example, among well-connected habitat patches, extinctions of local species may be avoided by the inflow of species from their surroundings. On the other hand, reduced connectivity caused by human-induced fragmentation (e.g., roads, dams) can have negative effects on population viability. However, the issue of connectivity and the maintenance of biodiversity should not be oversimplified. Under extreme conditions the probability of population survival for an overly connected system may be lower than that for a moderately connected system (Baggio et al., 2011; Salau et al., 2012).

In human social networks, connectivity can facilitate the resilience of ecosystem services through enhanced governance opportunities and functions. When high levels of connectivity between different social groups exist, information-sharing becomes smooth and the trust and reciprocity necessary for collective action can develop (Brondizio et al., 2009). However, just like in natural systems detrimental effects on the resilience of ecosystem services may develop when overconnectivity of social networks occurs.
**Related issues**

Though connectivity in general can facilitate recovery or constrain the spread of a disturbance, in certain situations it is also known to have detrimental effects on the resilience of ecosystem services. Depending on the nature and scale of disturbance, in a strongly connected SES the possibility of rapid propagation of disturbances through dense pathways between parts of the system exists, which may lead to widespread impacts on SES and associated ecosystem services (Van Nes and Scheffer, 2005; Ash and Newth, 2007). Typical examples of the high risk of propagation of disturbances in strongly connected systems can be found not only in ecological systems such as pest outbreaks and invasion of alien species, but also in socioeconomic systems such as financial crises, like the global spread of the 2008 recession triggered by the collapse of the US housing market (Adger et al., 2009; Biggs et al., 2011). In these cases, the resilience of ecosystem services, such as food production, pest regulation, and safety of the financial system, is negatively impacted by high connectivity. For more detailed discussions, refer to Dakos et al. (2015).

2.1.4 Principle 3—manage slow variables and feedbacks

**Definitions of slow variables and feedbacks**

SESs are self-organizing, CASs, which consist of and are influenced by a great number of variables that change and interact on a range of timescales. These variables can be divided into slow variables and fast variables, as some of them change much more gradually than others. Products of provisioning ecosystem services such as crop production and freshwater are examples of fast variables, and they are affected by soil composition and phosphorus concentration in lake sediments, which are examples of slow variables. Note that the notions of slow and fast variables are relative, and based on the context of the system each SES has its own timescales to determine its slow and fast variables (Walker et al., 2012).

In most SESs, it seems that a limited set of key variables and internal feedback processes interact to control the configuration of the system, and these key variables are usually slow variables. Thus, the underlying structure of SES is typically defined by slow variables, and the dynamics of the system arise from the interactions and feedbacks between fast variables under the conditions created by the slow variables (Gunderson and Holling, 2002; Norberg and Cumming, 2008). Regulating ecosystem services are often linked to slow ecological variables, such as erosion control, nutrient retention, and flood regulation (MA, 2003).

Feedbacks refer to the reversion effects on a variable, process, or signal under the current system conditions that have been influenced by an original change in that same variable, process or signal in an SES. The effect of feedbacks can either be reinforcing if it assists more change of the same kind, or damping if it resists further similar changes. These two feedbacks are also known as positive feedback and negative feedback. An example of reinforcing feedback is the exponential population growth with a positive growth rate: Population increase leads to more births
that increase the population. On the other hand, law enforcement measures against crime produce damping feedbacks by discouraging unlawful behavior.

**Resilience-enhancing mechanisms**

For all SESs, change and disturbance are assumed to be continuous threats, such as droughts and floods, while they are also exposed to more gradual, ongoing changes, such as increasing global trade connectivity, which often influence slow, controlling variables. It is important to know that the feedbacks within an SES contain critical information that reveals how the SES responds to such shocks and ongoing changes. In general, the possible configurations of an SES are shaped within certain bounds set by the controlling variables, which constitute the basic conditions for the system to develop its underlying structure and processes. For instance, under high rainfall conditions it is typical to find forests, and under low to medium rainfall conditions savannas (tropical grassland with scattered trees) are normally found. As the internal processes of an SES have little or only a very gradual influence on controlling variables, they are often regarded as being external to the system. Much research has been focused on understanding the links between controlling variables and SES outcomes.

Although in many situations, the configuration of an SES can be derived from the key controlling variables, this is not always the case. It is known that for a given set of conditions, it is sometimes possible for an SES to adopt two or more configurations with completely different sets of ecosystem services. For example, under intermediate rainfall conditions (around 1000–2500 mm per year), landscapes can exist either as open savannas (20%–40% tree cover) or as closed wooded landscapes (around 80% tree cover), with no substantial differences in soil and other factors (Sankaran et al., 2005; Hirota et al., 2011; Staver et al., 2011). In these situations, whether a particular place has savanna or forest at a particular point in time depends on the past configuration of the system, or more specifically, depends on which feedback processes have been dominant. With some disturbance, such as a drought or a large fire, the landscape can even abruptly shift from a forest to a savanna (or from a savanna to a forest with different disturbances), as a result of the change in the feedback processes triggered by the disturbance. These cases highlight the important role of the feedback processes in influencing SESs and the ecosystem services they provide.

In the face of change and disturbance, the self-organizing internal feedbacks of an SES typically buffer and absorb the impact of such shocks. Dampening feedbacks in particular are important in helping counteract disturbance and keep the system functioning consistently, thereby maintaining the current system configuration. In the previous example of imposing law enforcement measures against crime, the dampening feedbacks help keep social safety and maintain the way of life for ordinary people.

However, no SES can sustain infinite shocks or endure boundless changes without its basic framework and functions being fundamentally affected. All SESs have limits to how large a shock or how much change they can be exposed to and still recover and keep functioning as before. When critical limits in controlling slow
variables are exceeded, the feedbacks that keep the system in a particular configuration are significantly weakened or totally disappear, and thus the system is unable to buffer or absorb the changes. In the example of the forest to savanna shift, a key slow variable is the ratio of grass biomass to woody biomass, which decreases under the pressure of heavy grazing by cattle. When this ratio drops to a critical level there is too little grass to sustain fires that are sufficiently hot to burn and kill shrubs. As the small shrubs grow into trees, the SES will then shift to a forest configuration (Aneries et al., 2002). Note that in this case, fire is a key reinforcing feedback that maintains the grassy landscape of savannas.

As seen from this example, critical thresholds in controlling slow variables in an SES typically correspond to points at which previously dominant feedbacks disappear and new feedbacks suddenly come into operation, enforcing the SES to reorganize, often abruptly, into a different configuration. Consequently, the system is structured and functions in a different way, and produces a different set of ecosystem services. In ecological systems, such large, persistent and often abrupt system reorganizations are known as regime shifts. Regime shifts often occur suddenly, because the feedback mechanisms buffer change and leave little observable change to the system until the onset of regime shift. Similarly, changes in controlling slow variables, and thus the gradual erosion of resilience in the system, are frequently being ignored. Hence, a critical aspect of enhancing resilience in an SES is concerned with improving the management of slow variables and feedbacks to ensure that the system remains in a configuration that produces desired ecosystem services.

### Related issues

A well-known issue about feedbacks is that they can lock a system into an undesirable configuration or reduce its ability to adapt or transform in the face of change and disturbance. A typical example is the cycle of poverty in economics, in which reinforcing feedbacks keep the poor trapped in impoverished conditions that are hard to change. Other issues are mainly concerned with mismanagement of slow variables and feedbacks, which reduces the resilience of ecosystem services in SES. Examples include management interventions that obscure, remove, or ignore stabilizing feedbacks that underlie the provision of desired ecosystem services, or lack of knowledge about which key slow variables and feedbacks support particular SES configurations, or simply lack of action for improvement in regulating SES for various reasons. For more detailed discussions, refer to Biggs et al. (2015).

### 2.1.5 Principle 4—foster CAS thinking

#### Definitions of CAS thinking

Here, CAS thinking is defined as a mental model or worldview that considers SES as CAS and advocates taking relevant approaches to management accordingly. Based on this view, an SES is made up of many interacting components that are
adaptive to change, and hence the system is capable of self-organizing and evolving with new properties emerging at different scales and at different stages of development. Furthermore, a regime shift may take place in such a system, leading to a completely new SES that produces different ecosystem services.

As such, CAS thinking causes uncertainty in an SES as a system characteristic, making aspects of SES highly uncertain and difficult to predict and control. While active research and practical experiments have helped to reduce some important aspects of system uncertainty (Lee, 1993), CAS thinking regards some facets of uncertainty as irreducible due to unpredictability, incomplete knowledge or multiple knowledge frames (Levin, 2003; Brugnach et al., 2008), and embraces uncertainties as opportunities for positive motivational change (Janssen, 2002; Cilliers et al., 2013).

Note that fostering CAS thinking does not directly address the issue of enhancing the resilience of ecosystem services. Instead, it changes and adapts the cognitive foundations and paradigms that underpin management processes and decisions. In other words, recognizing that SESs are based on a complex and unpredictable web of connections and interdependencies is the first step toward management actions that can foster resilience in SES.

**Resilience-enhancing mechanisms**

Much of the theory that CAS thinking can help enhance the resilience of ecosystem services comes from the lessons learned from cases where conventional resource management with no CAS mindset has resulted in a loss of SES resilience. Based on many decades of widespread practices of ecosystem modification in the USA, Holling and Meffe (1996) described a “pathology of resource management” entailing practices such as river stabilization, fire suppression and monocultural farming to the point of system collapse. Such a pathology also characterizes mismanagement of agricultural activities (Allison and Hobbes, 2004), fisheries (Mahon et al., 2008), and forests (Agrawal, 2005), where an SES was narrowly managed to maximize economic production in the short term but this management approach degraded the underlying capacity of the system to produce ecosystem services. As a consequence groundwater tables were drawn down, land was degraded, fisheries, and forests were degraded or even depleted, and rivers were transformed and polluted.

These lessons show that management practices based on linear, reductionist worldviews of ecosystems inadvertently undermine the ability of these systems to continue producing ecosystem services in the face of disturbance and change. By inference, an alternative management approach based on a mindset that recognizes CAS properties may result in more resilient ecosystem services in the long term, because it considers consequences at a system level, across time, space, and actors (the resource users, managers, and policy-makers).

CAS thinking is not new, and cases exist where CAS thinking has contributed to improved socioecological outcomes through resilient ecosystem services. Recent examples of transformations in ecosystem management suggest that changes in
underlying mental models that acknowledge the characteristics of SES as CAS can lead to improvements in the resilience of ecosystem services. The large-scale rezoning of Australia’s Great Barrier Reef, one of the seven natural wonders of the world, is a case in point. This management change was driven by increased recognition of the importance of connectivity, nonlinear change and multiscale interactions in coral reef systems (Olsson et al., 2008). By introducing spatial restrictions on fishing and other uses of the reef, the rezoning aimed at enhancing the resilience of ecosystem functions to a range of disturbances including temperature anomalies and cyclones. CAS thinking in this approach is evident in two aspects: maintaining connectivity within the reef system and increasing the system’s capacity to absorb large disturbances; and importantly, recognizing the values and perspectives of different reef users. After decades of implementing this precautionary and adaptive management approach, improvements in the resilience of the reef’s marine ecosystems to climate-change impacts were evident through ecological monitoring and experimentation, but challenges remain as the problem domain has become increasingly complex (Brodie and Waterhouse, 2012; Brodie, 2014).

Despite examples that demonstrate the positive influence of CAS thinking on resilience building, due to the lack of solid scientific proof, it is still difficult to attribute the enhanced resilience directly to CAS mental models. This is not only due to the indirect influence of CAS thinking on management and resilience building, but also due to the short history of CAS thinking as a science with its application in management taking place only recently. In many cases more time is needed to assess the extent to which CAS thinking is guiding management or catalyzing positive change.

**Related issues**

Main issues related to fostering CAS thinking in management approaches to SES mainly result from misunderstanding of complexity, ineffective operations of new management approaches, and rigid interpretations of CAS worldviews. For instance, when “complexity” is interpreted merely as the unknown aspects of an SES, it may encourage managers to invest heavily in monitoring and data collection, rather than encourage the employment of adaptive approaches that allow for experimentation and the probing of boundaries as a mechanism to address uncertainty (Walters and Holling, 1990). Also, as a CAS approach implies a change in management paradigm from a focus on causality and control within short time-frames, to a focus on coping with change and uncertainty over longer timescales, ineffective operations in this management shift may occur. In addition, an incorrect attitude toward a CAS mental model may regard it as a static management form, rather than a dynamic management approach that pursues continued learning, experimentation and adaptation. As these problems may compromise resilience in SES to varying degrees, efforts should be made to avoid or solve them. For more detailed discussions, refer to Bohensky et al. (2015).
2.1.6 Principle 5—encourage learning

**Definition of learning**

According to Säljö (1979), learning means: (1) acquiring information and increasing knowledge; (2) memorizing; (3) acquiring facts, skills and methods; (4) making sense or abstracting meaning; and (5) interpreting and understanding reality in a different way by reinterpreting knowledge. Hence, learning can be understood as a multifaceted phenomenon. Two complementary approaches to learning are believed to enhance the resilience of ecosystem services, i.e., loop learning and social learning.

There are three types of loop learning: single-loop learning, double-loop learning, and triple-loop learning. Single-loop learning leads to a change in skills, practices or actions to meet existing goals and expectations, and it focuses on the question, “are we doing things right?” Double-loop learning involves questioning the assumptions that underlie action, and it focuses on the question, “are we doing the right things?” Triple-loop learning comprises a more in-depth questioning of values and norms that underlie institutions and actions, and it focuses on the question, “how do we know what the right thing to do is?” (Flood and Romm, 1996). By its nature, triple-loop learning can lead to the restructuring of beliefs and values, underlies transformations in worldviews, and may prompt changes in ecosystem governance and management approaches (Pahl-Wostl, 2009; Biggs et al., 2010).

According to Reed et al. (2010), social learning refers to “a change in understanding that goes beyond the individual to become situated within wider social units or communities of practice through social interactions between actors within social networks.” Social learning aims at sharing knowledge and perspectives, and takes place in two key ways, i.e., through deliberative processes of interactions between individuals, or through deliberate experimentation and reflection involving shared activities (Cundill and Rodela, 2012).

**Resilience-enhancing mechanisms**

As complex dynamical systems, SESs are constantly evolving and knowledge of these systems is always partial and becoming outdated. Enhancing the resilience of ecosystem services, therefore, requires continuous learning about the SES that provides these services (Holling, 1978; Walker and Salt, 2006; Chapin et al., 2009). Experience has shown that learning can enhance the resilience of ecosystem services primarily through influencing decision-making processes and governance. Such learning can be achieved through a variety of both planned and unplanned processes, in the forms of loop learning and social learning that include active experimentation and monitoring, multiactor collaboration and through intergenerational interactions with the environment.

Information about changes in the availability of ecosystem services can be obtained through experimentation and monitoring (Bellamy et al., 2001; Boyle
et al., 2001), which can also be used to resolve uncertainty about how the SES works. Also, the use of monitoring and experiments to explore alternative management options is an important means to support learning and enhance the resilience of ecosystem services.

There is growing recognition of the importance of broader participation in SES management and its role in enhancing the resilience of ecosystem services through learning (Danielsen et al., 2005). A study of five US community-based forestry organizations involved in collaborative ecological monitoring programs found that comonitoring activities led to single-loop learning, which identified changes in recommendations for optimal treatment of invasive weed species, and double-loop learning, which led to the realization of the significance and need for sustainable harvesting of mushrooms. These learning processes together altered social attitudes and assumptions toward adaptive management of SES. In the case of Australia’s Great Barrier Reef, the change of perceptions among politicians and the public from viewing the reef as pristine to viewing it as a severely threatened ecosystem cleared obstacles for stronger protection of the reef and its associated ecosystem services (Olsson et al., 2008). Both of these shifts in perceptions occurred through processes of learning. These experiences highlight the role of learning in potentially supporting the resilience of ecosystem services.

**Related issues**

There is growing recognition that learning can be ineffective, or worse, maladaptive to enhancing resilience in SES, and that the design of the learning process is crucial. A number of factors have been identified as potential challenges for effective and legitimate social learning and adaptive governance, which include both management issues and technical issues. The former influences how learning takes place, and the latter affects how effective monitoring and experimentation can be conducted and applied at the right scale. In general, to be effective, the process of monitoring, experimentation and learning should be collaborative and long-term, at an appropriate scale for decision-making and the SES, as well as able to withstand the impact of short-term funding cycles, politics and objectives (Barthel et al., 2010). Other aspects such as diverse participation, appropriate facilitation, sufficient financial and human resources and social networking are also considered to be important for effective learning. For more detailed discussions, refer to Cundill et al. (2015).

### 2.1.7 Principle 6—broaden participation

**Definition of participation**

According to Stringer et al. (2006), participation refers to the active engagement of relevant stakeholders in the management and governance process. Participation covers a wide range of engaging activities: from simply informing stakeholders to
complete devolution of management power. It can take place in all or some stages of an ecosystem services process: from identifying problems and goals to implementing policy and monitoring results, to assessing outcomes.

**Resilience-enhancing mechanisms**

As the interaction of the human race and nature is fundamental to the concept of SES, the role of participation in ecosystem management is well accepted (Schreiber et al., 2004; Armitage et al., 2007). Ample evidence suggests that, for ecosystem services to be resilient, a diversity of stakeholders should be involved in their management to improve legitimacy, to expand the depth and diversity of knowledge, and to help detect and assess system changes. Note that these elements are interdependent and interactive.

First, participation can improve legitimacy of SES management through establishing processes that are deliberative and support the formation or development of relationships between different stakeholders. These relationships create the potential to build trust and consensus as a basis for collective action around approaching thresholds, inventing innovative solutions or facilitating learning or shared experiences (Lebel et al., 2006).

Second, participation can also promote understanding of the system through enriched knowledge. Encouraging a variety of actors or stakeholders to participate can promote understanding of the SES dynamics by providing a range of ecological, social and political perspectives that may not be acquired through more traditional scientific processes (Armitage et al., 2009; Folke et al., 2005).

Third, participation can help to strengthen the link between information-gathering and decision-making, thus enabling responses to ecosystem change (Danielsen et al., 2005; Evans and Guariguata, 2008). Acting together, these mechanisms can enhance the capacity of a management system to detect and interpret shocks and disturbances, which is central to facilitating the collective action required to respond to changes in SES.

Again, the large-scale rezoning of Australia’s Great Barrier Reef is a case in point. An extensive public participation and consultation campaign was launched in an effort to raise awareness about threats to the reef, and to assist with new zoning plans to protect the reef (Olsson et al., 2008). Through greater understanding of the threats facing the reef, the public participation process was able to achieve both broad public support for the management decision on rezoning as well as the alteration of marine park zoning plans, in order to incorporate the concerns of some groups. In the related water governance sector, public participation was also found to broaden the range of interests considered in ecosystem-service management (Lebel et al., 2006).

**Related issues**

Although participation is essential to SES management, to actually enhance resilience in SES depends on factors such as the participants, the process and the social
and institutional environment (Stringer et al., 2006). These factors are interconnected and also context-dependent and, if not well prepared, executed, supported or resourced, participation can then undermine or compromise resilience.

The participants are crucial to the governance process because they determine or influence what ecosystem services are desired from a landscape and what should be the focus for resilience building. Resilience of ecosystem services can be compromised through participation that fails to engage the appropriate individuals and groups, or fails to consider the context-specific nature of the participation. Ineffective processes of participation can also be detrimental to ecosystem-service management, such as those that fail to have participatory strategies that successfully build social capital or fail to effectively link to natural systems. Participation processes may also compromise SES management if they fail to create a supportive social or institutional environment. Besides, attention should also be paid to balancing stakeholder interests, improving stakeholder relationships and choosing adequate scales of participation, because these factors can also influence the outcome of participation. For more detailed discussions, refer to Leitch et al. (2015).

2.1.8 Principle 7—promote polycentric governance systems

Definitions of polycentricity

Polycentricity refers to a governance system in which there are multiple interacting governing bodies with autonomy to make and enforce rules within a specific policy arena and geography. In an ideal polycentric system, to achieve a balance of collaboration and autonomy each individual governing body needs to interact and link with other authorities both horizontally and vertically. Here, governance is defined as the exercise of deliberation and decision-making among groups of people in the act of self-ordering their relationships. For instance, while national government agencies have the legitimate authority to make rules that are binding on all citizens, a regional management group (such as a forest or township management authority) may have autonomy to self-organize within its own domain.

Resilience-enhancing mechanisms

Since change and disturbance can occur at various possible levels ranging from global to regional and to local, polycentricity is believed to be a better governance system that enhances the resilience of ecosystem services in several fundamental ways than other monocentric governance strategies. These fundamental aspects all link with previously discussed resilience-enhancing principles.

First, a broadly inclusive system with governance at multiple smaller scales offers opportunities for experimentation at more localized levels, creating natural
experiments for testing different policies (Brondizio et al., 2009). Such processes encourage learning (Principle 5).

Second, by increasing the breadth of inclusion, a polycentric system can capitalize on scale-specific knowledge (e.g., traditional and local knowledge) to aid learning through sharing information, experience and knowledge across culture and scales (Olsson et al., 2004), while also expanding channels for managing slow variables and feedbacks (Principle 3) and providing opportunities for broader participation (Principle 6). Local levels with more direct linkages to the provision and use of resources provide institutional diversity from which successes can be shared with others (Folke et al., 1998). Furthermore, this decentralized design often serves to increase the legitimacy of the governance system at a given level with more scale-specific input into decision-making (Engle and Lemos, 2010), and also facilitates long-term monitoring and enforcement of rules at local levels, which are considered as important factors for building resilience in SESs.

Third, a polycentric approach has been suggested to confer connectivity and modularity (Principle 2), response diversity and functional redundancy (Principle 1), and foster CAS thinking (Principle 4) that can help preserve key SES elements in the face of disturbance and change. These characteristics of resilient governance may in turn change in scope and be further strengthened with increasingly collaborative degrees of polycentricity.

It should be pointed out, however, that studies to date in complex systems have been largely diagnostic and lacked predictive power and precision. To better understand how to operationalize, the idea of polycentricity in governance of SES requires more research.

**Related issues**

In polycentric governance, there are three key challenges that, if not resolved, may lead to degradation of ecosystem services at one or more scales. The first is that of the need to balance redundancy and experimentation, because overlapping authority and increasing transaction costs may result in inefficiencies in SES management (Parks and Ostrom, 1999). The second challenge is that of settling trade-offs between various ecosystem-service users (Rodriguez et al., 2006; Robards et al., 2011). Trade-offs may occur between conflicting goals and needs among users of current or potential ecosystem services (Søreng, 2006), or when impacts are incurred by those not affecting or benefiting from an ecosystem services (Chapin et al., 2006). The third challenge is closely related to the second one and is about politics, or the process of resolving conflict and making collective decisions over how to allocate trade-offs. Note that the issue of who bears the costs and who benefits from enhancing resilience in favor of particular ecosystem services constitutes one of the largest problems in SES governance (Lebel et al., 2006; Robards et al., 2011).

Finally it should be pointed out that, under some situations, particularly short timescales or crises where coordination across scales impedes timely action, there may be other governance tools (including top-down coercion or market approaches) that alone may accomplish specific goals more effectively than through a
polycentric system (Imperial and Yandle, 2005; Hilborn et al., 2006). For more detailed discussions, refer to Schoon et al. (2015).

2.1.9 Summary

This section discussed seven principles that enhance the resilience of ecosystem services in the face of change and disturbance. Although some principles are better understood than others, the evidence and resilience-enhancing mechanisms discussed show that all are important and require a good understanding of how, when and where they apply. Furthermore, as SESs are highly interconnected systems, the properties and processes associated with these principles often work together to become effective. Hence, context matters and promoting resilient ecosystem services depends as much on how the individual principles are applied as on achieving an appropriate combination of principles, keeping in mind their mutual interactions.

As interdependent CAS, the nature of SES calls for governance and management that enhances aspects of an SES that help shape future paths in favorable directions and enable adaptive responses to unexpected events. In essence, these principles help define the key features relevant for building resilience that should be considered when designing governance structures and management policies, which include:

1. Increasing understanding of critical SES components and processes as well as suitable management options;
2. Preparing the SES for unexpected events by creating awareness of their likelihood, and providing alternative approaches and ways of dealing with emergency issues when suddenly needed; and
3. Enhancing response capacity by providing a diversity of response options, building the trust needed to make decisions and take action, and providing ways to make use of different responses at the right scale.

More research is needed to better understand the individual principles, how they interact and how they can be operationalized and applied in different contexts.

References

Adger, W.N., Eakin, H., Winkels, A., 2009. Nested and teleconnected vulnerabilities to environmental change. Front. Ecol. Environ. 7, 150–157.
Agrawal, A., 2005. Environmentality: Technologies of Government and the Making of Subjects. Duke University Press, Durham, NC.
Allison, H.E., Hobbs, R.J., 2004. Resilience, adaptive capacity, and the ‘Lock-in Trap’ of the Western Australian agricultural region. Ecol. Soc. 9, 3.
Altieri, M.A., 2009. Agroecology, small farms, and food sovereignty. Mon. Rev. 61, 102–113.
Aneries, J.M., Janssen, M.A., Walker, B.H., 2002. Grazing management, resilience and the dynamics of a fire-driven rangeland. Ecosystems. 5, 23–44.
Books:

- Armitage, D., Berkes, F., Doubleday, N. (Eds.), 2007. Adaptive Co-Management: Collaboration, Learning, and Multi-Level Governance. UBC Press, Vancouver.
- Armitage, D., Plummer, R., Berkes, F., et al., 2009. Adaptive co-management for social—ecological complexity. Front. Ecol. Environ. 7, 95–102.

Articles:

- Ash, J., Newth, D., 2007. Optimising complex networks for resilience against cascading failure. Physica A. 380, 673–683.
- Baggio, J.A., Salau, K., Kanssen, M.A., Schoon, M.L., Bodin, Ø., 2011. Landscape connectivity and predator-prey population dynamics. Landsc. Ecol. 26, 33–45.
- Barnosky, A.D., Hadly, E.A., Bascompte, J., et al., 2012. Approaching a state shift in Earth’s biosphere. Nature. 486, 52–58.
- Barthel, S., Folke, C., Colding, J., 2010. Social—ecological memory in urban gardens: retaining the capacity for management of ecosystem services. Global Environ. Change. 20, 255–265.
- Bellamy, J., Walker, D., McDonald, G., Syme, G., 2001. A systems approach to the evaluation of natural resource management initiatives. J. Environ. Manage. 63, 407–423.
- Berkes, F., Folke, C. (Eds.), 1998. Linking Social and Ecological Systems. Cambridge University Press, Cambridge.
- Berkes, F., Colding, J., Folke, C. (Eds.), 2003. Navigating Social—Ecological Systems: Building Resilience for Complexity and Change. Cambridge University Press, Cambridge.
- Biggs, D., Biggs, R., Dakos, V., Scholes, R.J., Schoon, M.L., 2011. Are we entering an era of concatenated global crises? Ecol. Soc. 16, 27.
- Biggs, R., Gordon, L., Raudsepp-Hearne, C., Schlüter, M., Walker, B., 2015. Principle 3—manage slow variables and feedbacks. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge, pp. 105–141.
- Biggs, R., Schlüter, M., Schoon, M.L. (Eds.), 2015a. Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge.
- Biggs, R., Schlüter, M., Schoon, M.L., 2015b. An introduction to the resilience approach and principles to sustain ecosystem services in social—ecological systems. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge, pp. 1–31.
- Biggs, R., Westley, F., Carpenter, S., 2010. Navigating the back loop: fostering social innovation and transformation in ecosystem management. Ecol. Soc. 15, 9.
- Bohensky, E.L., Evans, L.S., Anderies, J.M., Biggs, D., Fabricius, C., 2015. Principle 4—foster complex adaptive systems thinking. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge, pp. 142–173.
- Boyle, M., Kay, J., Pond, B., 2001. Monitoring in support of policy: an adaptive ecosystem approach. In Encyclopedia of Global Environmental Change. Wiley, New York, NY, pp. 116–137.
- Brodie, J., 2014. Dredging the Great Barrier Reef: use and misuse of science. Estuarine Coastal Shelf Sci. 142, 1–3.
- Brodie, J., Waterhouse, J., 2012. A critical review of environmental management of the ‘not so Great’ Barrier Reef. Estuarine Coastal Shelf Sci. 104, 1–22.
- Brondizio, E.S., Ostrom, E., Young, O.R., 2009. Connectivity and the governance of multilevel social—ecological systems: the role of social capital. Annu. Rev. Environ. Resour. 34, 253–278.
Brugnach, M., Dewulf, A., Pahl-Wostl, C., Taillieu, T., 2008. Toward a relational concept of uncertainty: about knowing too little, knowing too differently, and accepting not to know. Ecol. Soc. 13, 30.

Chapin III, F.S., Lovecraft, A.L., Zavaleta, E.S., et al., 2006. Inaugural article: policy strategies to address sustainability of Alaskan boreal forests in response to a directionally changing climate. Proc. Natl. Acad. Sci. U.S.A. 103, 16637–16643.

Chapin, F.S., Kofinas, G.P., Folke, C. (Eds.), 2009. Principles of Ecosystem Stewardship: Resilience-Based Natural Resource Management in a Changing World. Springer, New York, NY.

CIA, 2013. The World Factbook 2013–14. Central Intelligence Agency, Washington, DC.

Cilliers, P., Biggs, H.C., Blignaut, S., et al., 2013. Complexity, modelling, and natural resource management. Ecol. Soc. 18, 1.

Cundill, G., Rodela, R., 2012. A review of assertions about the processes and outcomes of social learning in natural resource management. J. Environ. Manage. 113, 7–14.

Cundill, G., Leitch, A.M., Schulz, L., Armitage, D., Peterson, G., 2015. Principle 5—encourage learning. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge, pp. 174–200.

Dakos, V., Quinlan, A., Baggio, J.A., Bennett, E., Bodin, Ö., BurnSilver, S., 2015. Principle 2—manage connectivity. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge, pp. 80–104.

Danielsen, F., Burgess, N., Balmford, A., 2005. Monitoring matters: examining the potential of locally-based approaches. Biodivers. Conserv. 14, 2507–2542.

Elmqvist, T., Folke, C., Nyström, M., et al., 2003. Response diversity, ecosystem change, and resilience. Front. Ecol. Environ. 1, 488–494.

Engle, N.L., Lemos, M.C., 2010. Unpacking governance: building adaptive capacity to climate change of river basins in Brazil. Global Environ. Change. 20, 4–13.

Ernstson, H., 2013. The social production of ecosystem services: a framework for studying environmental justice and ecological complexity in urbanised landscapes. Landsc. Urban Plann. 109, 7–17.

Evans, K., Guariguata, M.R., 2008. Participatory Monitoring in Tropical Forest Management: A Review of Tolls, Concepts and Lessons Learned. CIFOR, Bogor.

Farman, J.C., Gardiner, B.G., Shanklin, J.D., 1985. Large losses of total ozone in Antarctica reveal seasonal C1O3/NO3 interaction. Nature. 315, 207–210.

Flood, R., Romm, N., 1996. Diversity Management: Triple Loop Learning. Wiley, Chichester.

Folke, C., Berkes, F., Colding, J., 1998. Ecological practices and social mechanisms for building resilience and sustainability. In Linking Social and Ecological Systems. Cambridge University Press, Cambridge, pp. 414–436.

Folke, C., Carpenter, S.R., Walker, B.H., et al., 2010. Resilience thinking: integrating resilience, adaptability and transformability. Ecol. Soc. 15, 20.

Folke, C., Hahn, T., Olsson, P., Norberg, J., 2005. Adaptive governance of social—ecological systems. Annu. Rev. Environ. Resour. 30, 441–473.

Gunderson, L.H., Holling, C.S. (Eds.), 2002. Panarchy: Understanding Transformations in Human and Natural Systems. Island Press, Washington, DC.

Hilborn, R., Arcese, P., Borner, M., et al., 2006. Effective enforcement in a conservation area. Science. 314, 1266.

Hirota, M., Holmgren, M., van Nes, E.H., Scheffer, M., 2011. Global resilience of tropical forest and savanna to critical transactions. Science. 334, 232–235.
Holling, C., 1978. Adaptive Environmental Assessment and Management. Wiley, London.
Holling, C.S., 1973. Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. 4, 1–23.
Holling, C.S., 1996. Engineering resilience versus ecological resilience. In: Schulze, Peter C. (Ed.), Engineering Within Ecological Constraints. National Academy of Engineering, National Academy Press, Washington, DC, pp. 31–44.
Holling, C.S., Meffe, G.K., 1996. Command and control and the pathology of natural resource management. Conserv. Biol. 10, 328–337.
Huntsinger, L., Oviedo, J.L., 2014. Ecosystem services are social—ecological services in a traditional pastoral system: the case of California’s Mediterranean rangelands. Ecol. Soc. 19, 8.
Imperial, M.T., Yandle, T., 2005. Taking institutions seriously: using the IAD framework to analyse fisheries policy. Soc. Nat. Resour. 18, 493–509.
IPCC, 2014. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge University Press, Cambridge.
Janssen, M.A., 2002. A future of surprises. In Panarchy: Understanding Transformations in Human and Natural Systems. Island Press, Washington, DC, pp. 241–260.
Janssen, M.A., Anderies, J.M., Ostrom, E., 2007. Robustness of social—ecological systems to spatial and temporal variability. Soc. Nat. Resour. 20, 307–322.
Jentoft, S., Bavinck, M., Johnson, D.S., Thomson, K.T., 2009. Fisheries co-management and legal pluralism: how an analytical problem becomes an institutional one. Hum. Organ. 68, 27–38.
Kotschy, K.A., 2013. Biodiversity, Redundancy and Resilience of Riparian Vegetation under Different Land Management Regimes. PhD Thesis. University of the Witwatersrand, Johannesburg.
Kotschy, K., Biggs, R., Daw, T., Folke, C., West, P., 2015. Principle 1—maintain diversity and redundancy. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge, pp. 50–79.
Lancaster, H.O., 1990. Expectations of Life: A Study in the Demography, Statistics, and History of World Mortality. Springer-Verlag, New York, NY.
Lebel, L., Anderies, J.M., Campbell, B.M., et al., 2006. Governance and the capacity to manage resilience in social—ecological systems. Ecol. Soc. 11, 19.
Lee, K.N., 1993. Compass and Gyroscope: Integrating Science and Politics for the Environment. Island Press, Washington, DC.
Leitch, A.M., Cundill, G., Schultz, L., Meek, C.L., 2015. Principle 6—broaden participation. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social—Ecological Systems. Cambridge University Press, Cambridge, pp. 201–225.
Leslie, P., McCabe, J.T., 2013. Response diversity and resilience in social—ecological systems. Curr. Anthropol. 54, 114–143.
Levin, S.A., 2003. Complex adaptive systems: exploring the known, the unknown and the unknowable. Bull. Am. Math. Soc. 40, 3–19.
MA, 2003. Ecosystems and Human Well-Being: A Framework for Assessment. Island Press, Washington, DC.
MA, 2005. Ecosystems and Human Well-Being: Synthesis. Island Press, Washington, DC.
Mahon, R., McConney, P., Roy, R.N., 2008. Governing fisheries as complex adaptive systems. Mar. Policy. 32, 104–112.
Martin, J., 2007. The Meaning of the 21st Century: A Vital Blueprint for Ensuring Our Future. Riverhead Books, New York, NY.
Mori, A., Furukawa, T., Sasaki, T., 2013. Response diversity determines the resilience to environmental change. Biol. Rev. 88, 349–364.

Mumby, P.J., Hastings, A., 2008. The impact of ecosystem connectivity on coral reef resilience. J. Appl. Ecol. 45, 854–862.

Norberg, J., Cumming, G.S. (Eds.), 2008. Complexity Theory for a Sustainable Future. Columbia University Press, New York, NY.

Nyström, M., Folke, C., 2001. Spatial resilience of coral reefs. Ecosystems. 4, 406–417.

Olsson, P., Folke, C., Berkes, F., 2004. Adaptive comanagement for building resilience in social–ecological systems. Environ. Manage. 34, 75–90.

Olsson, P., Folke, C., Hughes, T.P., 2008. Navigating the transition to ecosystem-based management of the Great Barrier Reef, Australia. Proc. Natl. Acad. Sci. U.S.A. 105, 9489–9494.

Ostrom, E., 2005. Understanding Institutional Diversity. Princeton University Press, Princeton, NJ.

Pahl-Wostl, C., 2009. A conceptual framework for analysing adaptive capacity and multilevel learning processes in resource governance regimes. Global Environ. Change. 19, 354–365.

Parks, R.B., Ostrom, E., 1999. Complex models of urban service systems. Polycentricity and Local Public Economies: Readings from the Workshop in Political Theory and Policy Analysis. University of Michigan Press, Michigan, MI, pp. 355–383.

Reed, M., Evely, A., Cundill, G., et al., 2010. What is social learning? Ecol. Soc. 15, 1.

Reyers, B., Biggs, R., Cumming, G.S., et al., 2013. Getting the measure of ecosystem services: a social–ecological approach. Front. Ecol. Environ. 11, 268–273.

Robards, M.D., Schoon, M.L., Meek, C.L., Engle, N.L., 2011. The importance of social drivers in the resilient provision of ecosystem services. Global Environ. Change. 21, 522–529.

Rockström, J., Steffen, W.L., Noone, K., et al., 2009. A safe operating space for humanity. Nature. 461, 472–475.

Rodriguez, J.P., Beard Jr, T.D., Bennett, E.M., et al., 2006. Trade-offs across space, time and ecosystem services. Ecol. Soc. 11, 28.

Salau, K., Schoon, M.L., Baggio, J.A., Janssen, M.A., 2012. Varying effects of connectivity and dispersal on interacting species dynamics. Ecol. Model. 242, 81–91.

Säljö, R., 1979. Learning in the learner’s perspective: some common-sense conceptions. Reports from the Institute of Education. University of Gothenburg, p. 76.

Sankaran, M., Hanan, N.P., Scholes, R.J., et al., 2005. Determinants of woody cover in African savannas. Nature. 438, 846–849.

Schoon, M.L., Robards, M.D., Meek, C.L., Galaz, V., 2015. Principle 7—promote polycentric governance systems. In: Biggs, Schlüter, Schoon (Eds.), Principles for Building Resilience: Sustaining Ecosystem Services in Social–Ecological Systems. Cambridge University Press, Cambridge, pp. 226–250.

Schreiber, E.S., Bearlin, A.R., Nicol, S.J., Todd, C.R., 2004. Adaptive management: a synthesis of current understanding and effective application. Ecol. Manage. Restor. 5, 177–182.

Søreng, S.U., 2006. Moral discourse in fisheries co-management: a case study of the Senja fishery, Northern Norway. Ocean Coast. Manage. 49, 147–163.

Staver, A.C., Archibald, S., Levin, S., 2011. The global extent and determinants of savanna and forest as alternative biome states. Science. 334, 230–232.

Steffen, W.L., Crutzen, P.J., McNeill, J.R., 2007. The anthropocene: are humans now overwhelming the great forces of Nature? AMBIO. 36, 614–621.
2.2 Resilience theories and practices in sociotechnical systems

Abbreviations

| Abbreviation | Description                                      |
|--------------|--------------------------------------------------|
| ATC          | air traffic control                              |
| EAS          | emergencies and abnormal situations              |
| FRMS         | fatigue risk management system                   |
| FSS          | financial services system                        |
| FTL          | flight and duty time limitation                   |
| STS          | sociotechnical system                             |
| T²EAM        | taskwork and teamwork strategies in emergencies  |

In this section, the resilience approach in sociotechnical systems (STSs) is introduced first, focusing on four main factors to build resilience in STSs. These factors represent four essential capabilities of a resilient system. Each of them is discussed through real case studies taken from various sociotechnical domains. For more detailed discussions, refer to the original publication (Hollnagel et al., 2011).
2.2.1 The resilience approach

Safety efforts are traditionally focused on the unwanted outcomes, injuries, and losses that are the result of adverse events. This fact is manifested by the common understanding of safety as “freedom from unacceptable risk.” The resilience approach, however, defines safety as the ability to succeed under varying conditions. As a result of this change in attitude toward safety, to learn from things that go right becomes equally important as to study things that go wrong. Hence, an understanding of the normal functioning of a STS provides the necessary and sufficient basis for understanding how it fails. Normally, it is both easier and more effective to improve safety by increasing the number of things that go right, than by reducing the number of things that go wrong.

The concept of resilience can be made more concrete by pointing to four abilities that are necessary for a system to be resilient. These are the ability to respond to events, to monitor ongoing developments, to anticipate future threats and opportunities, and to learn from past failures and successes alike; refer to Fig. 2.2. The engineering of resilience is concerned with establishing and managing these four capabilities. For STSs, resilience is defined as:

*The intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions.*

Note that the key term of this definition is the system’s ability to adjust its functioning. The following notes on the four essential capabilities of resilience give concrete meaning to the abstract concepts of this definition:

- Knowing what to do, that is, how to respond to regular and irregular disruptions and disturbances either by implementing a prepared set of responses or by adjusting normal functioning. This is the ability to address the actual.

![Figure 2.2 The four main capabilities of a resilient system.](image-url)
Knowing what to look for, that is, how to monitor that which is or can become a threat in the near term. The monitoring must cover both that which happens in the environment and that which happens in the system itself, that is, its own performance. This is the ability to address the critical.

Knowing what to expect, that is, how to anticipate developments, threats, and opportunities further into the future, such as potential changes, disruptions, pressures, and their consequences. This is the ability to address the potential.

Knowing what has happened, that is, how to learn from experience, in particular how to learn the right lessons from the right experience—successes as well as failures. This is the ability to address the factual.

To be able to respond to what happens in a timely and effective manner—whether it is a threat, or an opportunity—is essential for the survival of a system, organization, or organism. Responding means to detect and assess an event, and take relevant actions. To do so, it is necessary either to have prepared responses and the requisite resources, or to be flexible enough to make the necessary resources available when needed.

A resilient system must have effective means to monitor its own performance as well as changes in the environment. Monitoring enables the system to address possible near-term threats and opportunities before they actually take place. In order for the monitoring to be flexible, its basis must be assessed and revised frequently. While monitoring makes immediate sense, a specific effort should be made to look at the more distant future as well. The purpose of looking at the potential is to identify possible future events, conditions, or state changes that may affect the system’s ability to function either positively or negatively.

It is obvious that future performance can only be improved if something is learned from past performance. The effectiveness of learning depends on the basis for learning, that is, which events or experiences are taken into account, as well as on how the events are analyzed and understood. Since the number of things that go right, including near misses, is many orders of magnitude larger than the number of things that go wrong, it makes good sense to try to learn from representative events rather than from failures alone. For more detailed discussion, refer to Hollnagel (2011a).

2.2.2 Dealing with the actual: responding

Three Practical Cases Illustrating Resilience in “Real Time”

1. Lessons from the Hudson (Parié, 2011a)

On January 15, 2009, US Airways Flight 1549, a twin engine Airbus A320 made an unpwered emergency water landing on the Hudson River, in New York, after multiple bird strikes caused both engines to fail. All 155 passengers and crew aboard the Airbus A320 successfully evacuated from the partially submerged airframe as it sank into the river; they were rescued by nearby boats. The incident came to be known as the “Miracle on the Hudson,” which occurred when the plane turned into a flock of Canadian geese about two minutes after takeoff, at about 3000 ft (912 m). Several of these huge birds
were sucked into the twin engines, causing both engines to quickly lose power. As the aircraft lost altitude, the flight deck crew decided that the plane could not reach the closest airfield. They therefore diverted to the nearby Hudson River and executed a controlled water landing, about 3 minutes after losing power, saving all on board.

In this case study, the author discussed a “defense in depth” strategy against anticipated bird strikes by the aviation system, and the fact that this anticipation greatly contributed to survivability in the Hudson River event. This strategy is described as to lay three lines of defense: the first line in minimizing the frequency of bird strikes, the second line in assuring the ability of the aircraft and its engines to withstand some bird strikes, and the last line in enabling the crew-aircraft system to land or ditch safely after a total loss of power.

2. Coping with uncertainty: Resilient decisions in anesthesia (Cuvelier and Falzon, 2011)

In the context of research on patient safety in pediatric anesthesia, an empirical research was conducted in a pediatric anesthesiology service in a French hospital, based on the critical-incident technique (a set of procedures used for collecting direct observations of human behavior that have critical significance and meet methodically defined criteria). The objective of this study is twofold: to identify the different types of disturbances that anesthesiologists have to manage in their work, and to highlight the resilience factors, i.e., the strategies developed through practice by anesthesiologists to allow the system to function despite these disturbances.

The results highlight a distinction between potential situations in which the problem was envisioned beforehand by practitioners and unexpected situations which the anesthesiologists had not foreseen. This subjective classification based on “the astonishment of the perceiver” highlights two critical decisions made by anesthesiologists in order to manage variability. The first is the preoperative definition of an envelope of potential variability of the surgical intervention. The second concerns the occurrence of an event which trespasses that envelope and thus requires the mobilization of additional resources. The identification of these two critical decisions provides opportunities for research and actions to enhance resilience in the practice of anesthesia.

3. Training organizational resilience in escalating situations (Bergström et al., 2011)

To evaluate the possibility of building front-end organizational resilience in unexpected and escalating situations, an experiment was performed with Swedish fire safety engineers, using training programs that met methodically defined scenario guidelines. A 2-day crisis simulation exercise was used to practice the generic competencies of two experimental groups.

The experiment demonstrated that designing a training environment in which people actually faced the uncertainty and unpredictability of escalating situations generated “resilient” competencies that were not generated by current training strategies aiming at drilling correct behavior in known scenarios. It also showed that the nondomain specific training deepened understanding of the nature of escalating situations and the difficulties in managing them. The potential is great to apply this type of training in various industries that demand rapid and well-structured responses to escalating situations, although more research and further testing are needed.

Discussion

This section focuses on the ability of a system or an organization to “deal with the actual,” that is, to respond to the demands of the current situation—a disrupting or
shocking situation. At the “sharp end” of the system, “responding to the situation” implies a series of actions: assessing the situation, knowing what to respond to, and making decisions on what to do and when to do it. The readiness to respond mainly relies on two strategies: the proactive strategy and the reactive strategy. The first one is to anticipate the potential disruptive situations and predefine ready-for-use solutions (e.g., abnormal or emergency procedures, specific reaction skills, crisis response plans, and so on). The second is to generate, create, invent, or derive ad hoc solutions.

Put differently, the three cases discussed “real time” resilience, but from a synchronic as well as diachronic perspective. Indeed, from the viewpoints of designers, managers, or trainers, issues related to “real-time resilience” include how to ensure that the required resources (people, competence, equipment, etc.) are available or can be established in time. Hence, a more relevant question would be how to establish (now) and maintain (tomorrow) a readiness to respond (at any time in the future). While the three cases focus on the “readiness to respond,” they also touch on some issues related to establishing and maintaining that readiness. Each of them presents a practical case study of a specific domain: commercial aviation, anesthesia, and rescue services. Beyond their obvious differences of perspective and domain, they share similar underlying theoretical questions on resilience. The first of these is the relationship between resilience and anticipation, which also runs throughout other cases in this study on STS. For more detailed discussion, refer to Pariès (2011b).

### 2.2.3 Dealing with the critical: monitoring

**Three practical cases on monitoring for building system resilience**

1. From flight time limitations to fatigue risk management system (FRMS)—a way toward resilience (Cabon et al., 2011)

   Fatigue is known to be a major risk for safety in aviation. The flight and duty time limitations employed in civil aviation are a traditional approach to prevent fatigue through the regulation of duty hours. However, besides the inherent rigidity of regulations from an operational point of view, this method often fails to take into account all the complex dimensions of fatigue. In order to cope with this complexity, FRMSs are progressively emerging. Rather than setting absolute duty time limitations, an FRMS approach evaluates each operation in terms of fatigue risk. The International Civil Aviation Organisation (ICAO, 2011) defines an FRMS as “a data-driven means of continuously monitoring and maintaining fatigue-related safety risks, based upon scientific principles and knowledge as well as operational experience that aims to ensure relevant personnel are performing at adequate levels of alertness.”

   Monitoring plays an important role in FRMSs. Some suggestions on the overall monitoring process are given below, including four scenarios for developing a coherent strategy for monitoring fatigue risk:

   a. “Continuous” mode: In this “basic” mode, there is continuous feedback from the systematic monitoring on the risk matrix. For example, a risk that was not identified in the risk matrix is added after specific events are identified through air safety reports.
b. “Probe” mode: In this mode, a focused monitoring is conducted on a limited period (e.g., 1 month) and used to update the risk matrix.

c. Proactive mode: In this mode, a focused monitoring is triggered after a significant change (e.g., introduction of a new route, schedule change). The risk matrix is updated on the basis of the results.

d. Reactive mode: In this mode, a focused monitoring is triggered because of a significant change of an indicator of the systematic monitoring, e.g., frequency of air safety reports and associated aircrew fatigue report forms increased in the last months on a specific roster.

The implementation of FRMS requires a combination of tools and methods in order to manage the complexity of the impact of crew fatigue on their safety performance. FRMS can be seen as a concrete way to engineer resilience because it requires the organization to adjust its functioning by reintroducing safety managed by humans in addition to safety controlled by regulations.

2. Practices for noticing and dealing with the critical—a case study from the maintenance of power plants (Lay, 2011)

In high-risk, high-pressure complex work such as the maintenance of power plants, quality and safety incidents can happen and sometimes be extremely costly for both the service provider and the customer. Thus, consistent high-quality job performance with few incidents can be the most important differentiating factor when selecting a service provider. Conventional safety and quality programs are often limited in scope and tend to be micro-focused on specific, historical incidents or trends. Principles of resilience theories can be applied to design a broad, proactive strategy for noticing signs of a critical situation and moving into different actions to reduce the risk in order to improve performance in line with the plan.

Highly resilient organizations can be characterized by the following four behaviors:

a. They anticipate critical disruptions and situations and their consequences.

b. They notice the critical disruptions and situations when they occur.

c. They plan how to respond.

d. They adapt and move into different actions.

In this particular case, the first step in building resilience was improving “noticing” of the critical, looking at both general situations of an outage and specific unexpected situations. One approach was to implement a “pinging” process. Pinging is the proactive probing for risk profile changes (Wreathall and Merritt, 2003). Through workshops with experienced project managers and operational support staff, signs that an outage could be approaching an out-of-control situation or risk profile change were hypothesized. Another approach was to train the entire organization on concepts of resilience through study groups, by observing and recognizing the patterns and characteristics of resilience, and brittleness in every-day work, then discussing observations across situations and domains.

Field service is already seeing benefits from applying the concepts of resilience even though it is at the beginning of the journey to become more resilient. It seems that even just building the skill of “noticing” can help reduce loss. Noticing triggers action as people improvize responses even without a prescribed menu of help.

3. Cognitive strategies in emergency and abnormal situations training—implications for resilience in air traffic control (ATC) (Malakis and Kontogiannis, 2011)

The management of emergencies and abnormal situations (EAS) in the ATC system entails substantial challenges for air traffic controllers. The fundamental assumption behind refresher training is to provide air traffic controllers with the required skills and knowledge to meet successfully a wide range of challenges imposed by EAS. Using
cognitive systems engineering principles, a set of cognitive and team strategies was used to explore patterns of resilience in dyadic teams of operational controllers during real and simulated emergencies in a major European Area Control center. The investigation of real incidents revealed operational problems that were different from those encountered during refresher training. This case study presents initial findings with the potential of providing insights into cultivating sources of resilience that will supplement current refresher training.

The first stage of the study resulted in a performance model, termed “taskwork and teamwork strategies in emergencies in air traffic management (T2EAM).” The T2EAM model is an attempt to achieve a balanced and pragmatic approach to capture resilient processes during EAS episodes in the ATC system. The individual and joint cognitive strategies that correspond to T2EAM are listed below:

a. Anticipation: to timely and accurately detect and respond to a threat
b. Recognition: to timely and accurately detect early signs of an impending emergency and play out mentally the progression of events
c. Managing uncertainty: to assemble and assess a model of the situation and establish safety-related goals
d. Planning: to employ standard and/or contingency planning for the unfolding situation
e. Managing workload: to timely and accurately organize the required tasks and respond to interruptions and distractions
f. Team coordination and communication
g. Error management and task management

Based on the results of refresher training programs, it is concluded that these failure-sensitive cognitive strategies provide important practical examples of the potential for resilience at two levels. First, they provide insights on how adaptations by controllers in the form of cognitive strategies are employed to support resilience in the case of safety-critical events. Second, these cognitive strategies are used as building blocks in the development of advanced safety training programs with the aim of cultivating sources of resilience in the ATC system.

**Discussion**

Every organization concerned with safety uses one or more metrics to judge whether the levels of safety in the organization are acceptable. Using a common definition of safety, the organization needs to know if it is “free from unacceptable risk.” This metric is often the number or rate of accidents or injuries (or deaths) over some period of time or the time between events. While such measures may provide some assurance that safety has not been entirely out of control in the past, they are of little use when considering how to manage safety in the future and could even be harmful. Since the environment of the organization and its own internal processes are both dynamic, last year’s safety performance (or last month’s or yesterday’s) is at best a weak indication of how today’s and tomorrow’s will be.

As shown by the well-known management adage, “You can’t manage what you don’t measure,” measurement of processes is an essential part of any organization. Though in practice it is rare to find explicit models that provide a formal basis for identifying measures, knowledge often exists about relationships that are important to safety and that can be used to create working models. In the study of fatigue
(Case 1), a set of relationships among safety, fatigue, and its underlying mechanisms was discussed and those relationships were used to create metrics for monitoring the risks from fatigue. In the study of cognitive strategies in ATC (Case 3), the \( \text{T}^2 \text{EAM} \) performance model was developed based on existing research related to cognitive strategies in similar kinds of tasks, though often in other industries. This performance model provided the basis for developing measurement methods.

Several studies have used ad hoc methods for selecting indicators. As described in Case 2, the critical factors associated with an outage indicating a potential to be problematic were identified through brainstorming workshops with experienced project managers and others often involved in supporting projects when they get into trouble. In other words, even though these were based on learning from events in the past, they were considered to be sufficiently robust to remain effective for the next cycles of outages (1 to 2 years), at least from the perspective of testing the concept. Similarly, in work described by Wreathall (2006), the selection of indicators was based on factors most frequently associated with problems in human performance issues in the nuclear industry’s collective experience. For more detailed discussion, refer to Wreathall (2011).

### 2.2.4 Dealing with the potential: anticipating

**Three practical cases involving anticipation for building system resilience**

1. Measuring resilience in the planning of rail engineering work (Ferreira et al., 2011)

   The demand for increased capacity of the UK rail network has generated growing pressure to improve the planning and delivery of engineering work. As the owner of the UK rail infrastructure, Network Rail faces the challenge of delivering increasing volumes of work (maintenance, enhancements, and renewals) within more diverse and shorter opportunities for access to the infrastructure, while meeting the safety performance standards imposed by the regulatory bodies. Achieving a balance between productivity pressures and assurance of the required safety standards has become critical for the sustainability of the rail organization.

   Resilience engineering was proposed as a framework for research aiming to improve the ability of the organizational system responsible for the planning of all engineering work to respond to these pressures. Within this scope, an approach to measuring resilience was developed by means of questionnaire. A factor analysis method was used to identify underlying trends from the questionnaire data, which could then potentially be used as measurable aspects of resilience in rail engineering planning. The identified resilience factors are listed below:

   a. Adaptability and flexibility: Planners are able to restructure their work in response to pressures and adapt to new arising circumstances through problem solving.

   b. Control: People feel they have the means necessary, in particular information, to appropriately control and steer their activities.

   c. Awareness and preparedness: The system generates feedback and provides support in such a way that people have a clear view of how they should contribute to responding to challenges.
d. Trade-offs: Achieving a balance between safety and efficiency through decision-making.
e. Time management: Having the time to be thorough when required by planning decisions.

Self-reporting methods, such as questionnaires, may be insufficient to provide a robust measure of resilience. However, such methods can be an efficient way of monitoring the system’s behavior in terms of resilience, particularly on a longer term basis through periodical applications.

2. The art of balancing: using upward resilience traits to deal with conflicting goals (Tjørhom and Aase, 2011)

This case study describes some of the processes involved in balancing conflicting goals (e.g., between safety and operation) in a change-intensive environment by using examples of the Norwegian civil aviation transport. The ability to handle multiple goals involves the use of both downward and upward resilience traits to address potential conflicts. Here, downward resilience means that macrolevel directions and solutions prepare for resilience through clear goal structures, infrastructure, and procedures that handle the trade-offs between safety and efficiency. Upward resilience means that decisions made at the microlevel in a system reflect a commitment to safety in case of goal conflicts. Changes, caused either by external or internal drivers, may alter these resilience traits by introducing loss of oversight. Changes made at the macrolevel of the system might have unintended consequences on the microlevel, and vice versa.

These findings have implications for different levels of the aviation transport system. The following actions are proposed to strengthen downward resilience:

a. Development of clear safety goal rules at the governmental level.
b. Downward resilience is threatened by unwillingness to state clear goal rules at the strategic level. After years of changes within the aviation transport system, employees need clear statements that give them a framework to remain flexible and committed to safety despite economic pressure.
c. The goal rules should be based on worst-case scenarios using input from the entire aviation transport system. The institutional level of the system must be responsible for collecting information regarding trends that threaten resilience.
d. Development of guidelines and requirements for addressing crossscale interactions.
e. The training tools should include participants from different levels and professions.

The following actions are proposed to strengthen upward resilience:

a. Foster perpetual awareness among operators.
b. Without a constant sense of unease about the way to handle an operation, one might become lost in routine and fail to notice variations. Even a seemingly insignificant variance in operation must be taken as a potential indicator of a threat to resilience.
c. Extend operators’ collaboration with other parts of the system.
d. A strong focus on professional values might have some downsides (McDonald, 2006). Within a profession, self-confidence may evolve to the level of overconfidence. In a trade-off situation, this may result in over-reliance on the individual’s judgment—at the expense of cautious prudence. Technicians and airport operators might rely too heavily on experience and knowledge, thus taking unnecessary chances without fully embracing the body of rules. Interrelations necessitate an exchange of knowledge across professions.

The tension between downward and upward resilience in the aviation system studied is balanced by a strong professionalism throughout the system, which functions as a buffer and makes safety goals prevalent over production goals. To uphold this art of balancing,
it is crucial to develop strong but flexible goal rules at the macrolevel to demonstrate a commitment to safety that microlevel actors find trustworthy. At points of intensified production pressure and higher organizational tempo, extra investments in sources of resilience are required to keep production/safety trade-offs from sliding out-of-balance. In other words, safety investments are most important when least affordable (Woods, 2006).

3. The importance of functional interdependencies in financial services systems (FSSs) (Sundström and Hollnagel, 2011)

The events of 2007–2009 in the global financial markets clearly illustrated the need for an improved understanding of how the global FSS functions. In particular, the crisis made it clear that national FSSs, or components of such systems such as individual banks, were highly dependent on the normal functioning of other components of the global FSS.

The primary purpose of this case study was to illustrate how resilience engineering can provide the financial services industry with a different way to understand risk at both a macro- and micropurudential level. To enable multiple stakeholders to have a shared view of the system, a common model of the FSS is required. In the present study, the functional approach to FSSs proposed by Merton and Bodie (1995) was employed. Such an approach enables stakeholders and decision makers to focus on behavior rather than on the specifics of an individual financial services institution. The modeling process typically consists of four phases:

a. Identify what functions need to be modeled.
b. Identify conditions that could lead to a change in performance.
c. Identify areas where functional resonance could emerge.
d. Identify how performance variance can be monitored and controlled.

A key advantage of a functional perspective is that it becomes possible to discover risks created by functional interdependencies among individual institutions and system components. The concepts and functional modeling approach outlined in this study thus provide the basis for the development of a standardized method to capture and better understand risk in the financial services industry.

Discussion

The ability to anticipate and adapt runs throughout the discussions of resilience in this section. The case studies identified several patterns in how resilient systems may anticipate that adaptive capacity is falling, that buffers or reserves may become exhausted, that goal priorities should be changed, etc. These patterns include:

- Resilient systems are able to recognize that adaptive capacity is falling.
- Resilient systems are able to recognize that buffers or reserves become exhausted.
- Resilient systems are able to recognize when to shift priorities across goal trade-offs.
- Resilient systems are able to make perspective shifts and contrast diverse perspectives that go beyond their nominal position.
- Resilient systems are able to navigate changing interdependencies across roles, activities, levels, and goals.
- Resilient systems are able to recognize the need to learn new ways to adapt.

To be resilient, a system always keeps an eye on whether its adaptive capacity, as it is configured and performs currently, is adequate to meet the demands it may encounter in the future. Overlooking or discounting signs that adaptive capacity is
degrading leaves that system vulnerable to sudden collapse or failures (Woods, 2009). For more detailed discussion, refer to Woods (2011).

### 2.2.5 Dealing with the factual: learning

The following views on learning are adapted from Hollnagel’s work, “To learn or not to learn, that is the question” (Hollnagel, 2011b).

#### The conditions for learning

In order for learning to take place, three conditions must be fulfilled. The first condition is that there are reasonable opportunities to learn, that is, that situations where something can be learned occur with a sufficiently high frequency. The second condition is that the situations are sufficiently similar to allow generalizations to be made, especially regarding their reasons or causes. The third condition is that there must be sufficient opportunity to verify that the right lessons have been learned.

When accidents, emergencies, and catastrophes occur, it is clearly important to find out why they happened but it is also clear that they do not offer the best basis for learning. Accidents do not happen very frequently, at least if the domain of activity is reasonably safe. Furthermore, accidents are usually different from each other, and the differences are often proportional to the magnitude of the outcomes. Lastly, because accidents happen so rarely, there is little opportunity to check whether the right lessons have been learned. Accidents therefore do not provide good conditions for learning, common stereotypes notwithstanding.

It follows from these arguments that learning can be more effective if it is based on events or conditions that happen more frequently and that—almost by virtue of that fact—are less extreme and less dissimilar (e.g., Herrera et al., 2009; Woods and Sarter, 2000). Indeed, it is more efficient to learn from what goes right than to learn from what goes wrong, because the former happens far more often than the latter.

#### The impact of learning

The four main capabilities of resilience are equally necessary and therefore equally important. Taking learning as a starting point, it is easy to argue that the ability to respond would be of little value without the ability to learn. Facing change and disturbance, it is imperative to learn new ways to respond and the system can learn only by observing and evaluating the efficiency of the responses.

A similar kind of argument can be made for the relation between learning and monitoring. It is mainly by learning through practice that the proper basis for monitoring—the indicators that must be watched—can be established. The efficiency of monitoring depends on the efficiency of learning, just as the efficiency of learning depends on looking at the right kinds of experiences. Finally, learning is also
necessary for anticipation, which is essential to produce a realistic, or even adequate, model or understanding of what may possibly happen in the future.

### 2.2.6 What is the nature of resilience? (McDonald, 2006)

If resilience is a system property, then it probably needs to be seen as an aspect of the relationship between a particular STS and the environment of that system. Resilience appears to convey the properties of being adapted to the requirements of the environment, or otherwise being able to manage the variability or challenging circumstances the environment throws up. An essential characteristic is to maintain stability and integrity of core processes, despite perturbation.

The focus is on medium- to long-term survival rather than short-term adjustment per se. However, the organization’s capacity to adapt and hence to survive becomes one of the central questions about resilience—because the stability of the environment cannot be taken for granted. Therefore, the notion of being able to read the environment appropriately and to be able to anticipate, plan, and implement appropriate adjustments to address perceived future requirements, is important.

Resilience represents the capacity (of an organizational system) to anticipate and manage risk effectively, through appropriate adaptation of its actions, systems, and processes, so as to ensure that its core functions are carried out in a stable and effective relationship with the environment.

### References

Bergström, J., Dahlström, N., Dekker, S., Petersen, K., 2011. Training organisational resilience in escalating situations. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 45–57.

Cabon, P., Deharvengt, S., Berechet, I., Grau, J.Y., Maille, N., Mollard, R., 2011. From flight time limitations to fatigue risk management systems—a way toward resilience. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 69–86.

Cuvelier, L., Falzon, P., 2011. Coping with uncertainty: resilient decisions in anaesthesia. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 29–43.

Ferreira, P., Wilson, J.R., Ryan, B., Sharples, S., 2011. Measuring resilience in the planning of rail engineering work. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 145–156.

Herrera, L.A., Norsdkag, A.O., Myhre, G., Halvorsen, K., 2009. Aviation safety and maintenance under major organisational changes, investigating non-existing accidents. Accident Analysis and Prevention. 41 (6), 1155–1163.

Hollnagel, E., 2011a. Prologue: the scope of resilience engineering. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA.
Hollnagel, E., 2011b. To learn or not to learn, that is the question. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 193–198.

Hollnagel, E., Pariés, J., Woods, D., Wreathall, J. (Eds.), 2011. Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA.

ICAO, 2011. FRMS Implementation Guide for Operators. first ed. IATA, ICAO, IFALPA.

Lay, E., 2011. Practices for noticing and dealing with the critical: a case study from maintenance of power plants. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 87–100.

Malakis, S., Kontogiannis, T., 2011. Cognitive strategies in emergency and abnormal situations training—implications for resilience in air traffic control. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 101–117.

McDonald, N., 2006. Organisational resilience and industrial risk. In: Hollnagel, E., Woods, D.D., Leveson, N. (Eds.), Resilience Engineering: Concepts and Precepts. Ashgate, Aldershot, UK, pp. 155–180.

Merton, R.C., Bodie, Z., 1995. A conceptual framework for analysing the financial environment. In: Crane, D.B., Froot, K.A., Mason, S.P., Perold, A.F., Merton, R.C. (Eds.), The Global Financial System. Harvard Business School Press, Cambridge, MA, pp. 3–31.

Pariés, J., 2011a. Lessons from the Hudson. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 9–27.

Pariés, J., 2011b. Resilience and the ability to respond. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 4–8.

Sundström, G., Hollnagel, E., 2011. The importance of functional interdependencies in financial services systems. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 171–190.

Tjørhom, B., Aase, K., 2011. The art of balance: using upward resilience traits to deal with conflicting goals. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 157–170.

Woods, D.D., 2006. How to design a safety organisation: test case for resilience engineering. In: Hollnagel, E., Woods, D.D., Leveson, N. (Eds.), Resilience Engineering: Concepts and Precepts. Ashgate, Aldershot, UK, pp. 315–325.

Woods, D.D., 2009. Escaping failures of foresight. Saf. Sci. 47 (4), 498–501.

Woods, D.D., 2011. Resilience and the ability to anticipate. In: Hollnagel, E., Pariés, J., Woods, D.D., Wreathall, J. (Eds.), Resilience Engineering in Practice: A Guidebook. Ashgate Publishing Company, Burlington, USA, pp. 121–125.

Woods, D.D., Sarter, N.B., 2000. Learning from automation surprises and ‘going sour’ accidents. In: Sarter, N., Amalberti, R. (Eds.), Cognitive Engineering in the Aviation Domain. Erlbaum, Hillsdale, NJ, pp. 327–354.

Wreathall, J., 2006. Properties of resilient organisations: an initial view. In: Hollnagel, E., Woods, D.D., Leveson, N. (Eds.), Resilience Engineering: Concepts and Precepts. Ashgate, Aldershot, UK, pp. 275–285.
2.3 Resilience theories and practices in computer systems

Abbreviations

| Abbreviation | Description                     |
|--------------|----------------------------------|
| CPU          | central processing unit         |
| DDD          | displacement damage dose        |
| DRE          | detected recoverable error      |
| DUE          | detected unrecoverable error     |
| FM           | fault model                     |
| FT           | fault tolerance                 |
| GAFT         | generalized algorithm of fault tolerance |
| HW           | hardware                        |
| IC           | integrated circuit              |
| NMR          | N-modular redundancy            |
| PKA          | primary knock-on atom           |
| RTS          | real-time system                |
| SDC          | silent data corruption          |
| SEE          | single event effect             |
| SSW          | system software                 |
| TID          | total ionizing dose             |
| TMR          | triple modular redundancy       |
| TR           | time redundancy                 |

In this section, a paradigm for designing new-generation resilient and evolving computer systems is presented, including its key concepts, elements of supportive theories, methods of analyses and implementation principles. For more detailed discussions, refer to the original publication (Castano and Schagaev, 2015).

2.3.1 Emerging challenges

Most embedded systems, in which the computer is completely encapsulated by the device it controls and performs predefined tasks, are real-time systems (RTSs). Since their early applications in the 1960s, embedded RTSs have been woven into the fabric of modern society in ubiquitous ways, built into homes, offices, bridges,
medical instruments, cars, airplanes and satellites, and even into clothes. They are being used in fields where their correct operation is vital to ensure the safety and security of the public and the environment: from automotive systems and avionics to intensive health care and industrial control as well as military operations and defense systems. There are time constraints imposed on these systems for responses, and their safety-critical nature demands the highest possible availability and reliability of system operation.

The technological achievements that have led to exponential growth of clock frequency (at which a chip like a central processing unit is running) and memory size have led to the development of microprocessors with increased transistor density. This growth has been made possible by the progressive miniaturization of electronic components predicted by Moore, whose 1965 paper described a doubling every year in the number of components per integrated circuit (IC) (Moore, 1965). Eventually, technological development has been slowed by physical limitations: due to the size reduction of electronic components to nanometer scales and due to the increase in clock frequencies, supply voltages have been reduced to keep power dissipation manageable but thermal noise voltages have increased (Asanovic et al., 2006; Kish, 2002).

Environmental impact and radiation effects have long been a serious concern in aviation, aerospace and special mission electronics. Radiation-induced faults are frequent in outer space (Adams et al., 1982; Binder et al., 1975; Blake and Mandel, 1986). Due to the reductions in size and voltage, the sensitivity of embedded systems to ionizing particles has increased considerably. Energizing particles can produce a number of faults at the hardware (HW) level, not only in harsh environments such as outer space but also in real-life environments, which has been verified by stress experiments in the laboratory using particle bombardment.

Hence, electronic components with lower power and noise margins are less reliable, which is why recent systems are more prone to transient faults induced primarily by radiation (Baumann, 2002, 2005a,b; Seifert et al., 2002; Shivakumar et al., 2002). Although transient faults do not cause permanent damage in circuits, they can affect system behavior by corrupting stored information or signal communication (Karnik and Hazucha, 2004; Mavis and Eaton, 2002).

In view of these new challenges, there is an increasing need to deal with faults and their consequences in the system. There are two classes of mechanism to cope with faults: fault avoidance and fault tolerance (FT) (Avizienis et al., 2004). Fault avoidance means developing components/systems that are less likely to present faults, while fault-tolerant techniques enable the system to tolerate the effects of these faults. By definition, FT is the ability to provide uninterrupted services, conforming to the desired levels of reliability even in the presence of faults (Avizienis et al., 2004). As complete avoidance of faults in a system is practically impossible, a balance of the two approaches is currently applied.

Despite the efforts made to apply fault-tolerant techniques to commercial off-the-shelf computers to reduce costs, no substantial breakthrough has been achieved in efficiency or any other crucial property required of safety-critical systems (Antola et al., 1986). The main focus of the research community is (1) identifying all possible mechanisms leading to accidents and (2) providing preplanned defense
techniques against them. However, too little attention has been paid to developing potentially resilient systems that can withstand deviations from desired states due to change and disturbance.

This research on new-generation resilient and evolving computer systems is driven by observations on resultant limitations from the evolution of computer architecture, which have been motivated by technological and market choices as well as physical limitations. As the traditional concepts of reliability, FT and dependability (the ability to deliver a service that can justifiably be trusted) do not take into account the transient nature of some of the faults induced by radiation, a new concept of resilience has to be introduced for embedded systems, reflecting the changing environment and the different FT contexts. The proposed resilience theory possesses six attributes: reliability, security, safety, performability, robustness, and evolvability. Note that software faults as the source of errors are out of the scope of this work.

2.3.2 The resilience approach

In computer science the term “resilience” has been traditionally used as a synonym of FT. Historically, the word has been used in various fields with multiple meanings. As a property, it has different implications. In social psychology resilience is about elasticity, spirit, resource and good mood. On the other hand, in material science resilience involves not only elasticity but also robustness. In the present study, the concept of resilience for safety-critical applications is extended in line with the material science usage. Therefore, the term resilience discussed in this section includes both attributes: robustness and elasticity.

The term “robustness” refers to the use of static techniques such as the use of very reliable materials or the use of rigid and predesign approaches to FT. Ideally, a robust system can deliver correct service in conditions beyond the normal domain of operation without fundamental changes to the original system. However, total reliability in the event of unforeseen faults other than the normal domain of operation is not feasible. Therefore, a resilient approach requires another important attribute, elasticity, to respond to deviations from desired states due to change and disturbance.

Elasticity is interpreted as the ability to “spring back” (or recover) without losing the intrinsic properties of the material or system. Applied to resilience, elasticity is understood as the ability to evolve, to successfully accommodate changes (evolvability). An evolvable system may perform changes to the system, decreasing its level of performance or reliability for a specific time range (1) to compensate for faults or (2) during exceptional circumstances (graceful degradation). In other words, a resilient system must have the ability to be adaptable, understanding adaptability as the ability to evolve while executing. Hence, adaptability is a subset of evolvability and requires the ability to anticipate changes prior to the occurrence of the resulting damage.
Therefore, a resilient architecture must possess different mechanisms to acquire both attributes: (1) static predesign fault-tolerant techniques (robust) and (2) dynamic techniques (elastic) that may be achieved with the ability to reconfigure elements of the system (reconfiguration). Such a resilient approach in designing the computer architecture for safety-critical applications aims at achieving the ability to deliver correct service adapting to disturbance, disruption, and change within specified time constraints.

A list of specific requirements or attributes is given as follows:

- Continuity of service (reliability)
- Readiness for usage (availability)
- Nonoccurrence of catastrophic consequences (safety)
- Nonoccurrence of incorrect system alterations (integrity)
- Ability to undergo corrective maintenance and recovery with maximum coverage of faults (testability and recoverability)
- Ability to perform in the presence of faults (performability)
- Ability to decrease the level of performance for a specific time range in order to compensate for HW faults (graceful degradation)
- Ability to regain operational status via reconfiguration in the presence of faults (recoverability via reconfiguration)
- Ability to accommodate changes (evolvability)
- Ability to anticipate changes (adaptability)

Clearly, resilience is not a simple single concept, but is a function of all these attributes. Taking into consideration all these attributes, resilience is defined as follows:

A resilient system is a system that over a specified time interval, under specified environmental and operating conditions, is ready to perform its intended function, guaranteeing the absence of improper system alterations, having the ability to anticipate and accommodate changes while executing, and the ability to conduct servicing and inspections so that in case of failure quick restoration to a specified working condition can be achieved, or otherwise discontinuation of operation in a safe way is performed.

Fig. 2.3 illustrates the different attributes and measures of resilience. As seen, most of them are important design principles of traditional computer systems. For clarity, the definitions of reliability, safety, security, and performability as discussed here are given below:

Reliability: Reliability R(t) is the probability that a system or component will perform its intended function (without failure) over the entire interval [1, t] under specified environmental and operating conditions.

Safety: In safety-critical systems, safety describes the absence of catastrophic failures for users and the environment when a failure takes place.

Security: Security is the concurrent existence of three attributes: integrity, maintainability and availability. Integrity can be defined as the absence of improper system state alterations. Maintainability refers to the ease and rapidity with which, following a failure, a repairable system can be restored to a specific operational condition. Availability can be simply defined as “readiness for correct service” (Avizienis et al., 2004).
Performability: Performability is the ability of a system or component to accomplish its designated functions within specified constraints such as speed, accuracy or memory usage.

2.3.3 Impact of radiation on electronics

Hardware faults are a major concern in silicon-based electronic components such as static random access memory, dynamic random access memory, microprocessors and field programmable gate arrays, which have a well-documented history of faults mainly caused by high-energy nuclear particles. In the case of safety-critical systems, maximum reliability can be achieved by assuming susceptibility of those systems to faults produced by various internal (e.g., interconnect coupling noise) and external (e.g., cosmic and solar radiation) disturbances. The traditional reliability analyses of these systems assume failure rates of permanent faults. So far, design and reliability engineers have discounted the effect of transient faults. However, the density of modern silicon chips makes them vulnerable to particles of lower energy that can cause transient faults and consequently, catastrophic system failures (Constantinescu, 2003; Hazucha and Svensson, 2000; Hazucha et al., 2003).

Radiation and its effect on electronics

The term “radiation” commonly refers to the process in which energy in the form of waves or particles travels through space or a material medium, ultimately to be absorbed by another body. In general, radiation can be divided into ionizing and nonionizing radiation depending on its ability to ionize matter. Nonionizing radiation is not a concern here because normally it does not carry enough energy to produce changes to electronic circuitry.
On the other hand, ionizing radiation carries enough energy to directly or indirectly remove electrons from atoms or molecules, thus causing the formation of ions. Sources of ionizing radiation include highly energetic protons, alpha particles, heavy ions, galactic cosmic rays, and others. This classification also includes neutrons because their collision with nuclei produces ionizing radiation, even though they are not ionizing particles.

Due to the reduction in size of the transistors and the reduction in critical charge of logic circuits, the effects of ionizing radiation are becoming a serious concern in semiconductor devices in various settings, such as outer space, high altitudes and at sea level. Basic radiation damage mechanisms affecting electronics include atomic lattice displacements and ionization damage, which induce different types of failure such as total ionizing dose (TID), single event effects (SEEs), and displacement damage dose (DDD). Thus, when energetic particles collide with sensitive regions of the semiconductor, stored information can be distorted, potentially leading to logic errors. Ironically, as manufacturing technologies advance, the natural resilience of previous technologies to information corruption is decreasing (Baumann, 2002, 2005a,b; Seifert et al., 2002; Shivakumar et al., 2002).

Transient faults (Breuer, 1973) are the predominant faults in modern technologies, and can be caused by environmental conditions like temperature, pressure, humidity, voltage, power supply, vibrations, fluctuations, and electromagnetic interference due to crosstalk between long parallel lines in a chip. However, ionizing particles are the major source of this type of fault.

**Damage mechanisms**

As mentioned above, there are two basic damage mechanisms to electronic elements due to radiation: atomic lattice displacement and ionization. Atomic lattice displacement occurs when an energetic particle undergoes a nuclear collision with one or more atoms of the electronic device, changing its original position (refer to Fig. 2.4) and thus the analog properties of the semiconductor junctions, potentially affecting the properties of the material and creating lasting damage. As shown in Fig. 2.4, the displaced atom is referred to as “primary knock-on atom” (PKA), and its new nonlattice position is called “interstitial,” while its absence from its original lattice position is named “vacancy.” The combination of a vacancy and an adjacent interstitial is called a “Frenkel pair.”

In silicon, displacement of an impacted atom is possible if it is part of the crystalline structure and the incident particle is capable of inducing a minimum energy (displacement threshold energy) of around 20 eV (Miller et al., 1994). In most cases, the displaced atom with enough energy is likely to further knock out a neighboring atom creating a more complex configuration known as a “cluster,” affecting the properties of the bulk semiconductor material. In silicon, clusters and vacancies are of unstable nature and tend to be filled by nearby atoms, forming more stable defects. In general, this migration leads to the most typical process, known as “defect reordering” or “forward annealing,” mitigating the amount of damage and
its effectiveness. Yet, in some cases, depending on the time, temperature and nature
of the device, “reverse annealing” can also occur, producing more efficient defects.

It is known that ionization damage is primarily induced by charged particles,
usually leading to transient effects that cause temporary variation of the functional-
ity of the system. This type of error is called a soft error because no permanent
damage is induced in the electronic circuit. Ionization damage may also lead to
small degradation and permanent errors, which are categorized as hard errors.

A key factor in the damage process is the critical charge, or $Q_{\text{crit}}$, which repre-
sents the smallest amount of charge that can cause a change of value in a cell. The
effects provoked by the above damage mechanisms can vary depending on the type
of radiation, radiation flux, total dose, critical charge of the device and manufactur-
ing technology, making fault modeling difficult, and time consuming.

Radiation macro effects

The resultant macro effects of radiation may be classified into three categories:
TID, DDD, and SEEs. Based on the type of degradation that these macro effects
generate on electronics, TID and DDD are considered as long term and SEEs as
short term.

TID is a measure of the cumulative effects of prolonged exposure to ionizing
radiation. Metal oxide semiconductor and bipolar electronic technologies are
affected by TID and the resulting damage is permanent: once the material is dam-
aged, it will not return to its original state (Felix et al., 2007). Typical TID effects
include parametric failures, or variations in device parameters such as leakage cur-
rent, threshold voltage, etc., or functional failures. The primary sources of TID are
trapped protons and electrons, and solar protons (Barth et al., 2004).

DDD or “bulk” damage (Barth et al., 2004; Yu et al., 2005) occurs when high-
energy particles displace atoms from the semiconductor lattice due to its prolonged
exposure to nonionizing energy loss. DDD often has similar long-term degradation
characteristics to TID, but is a separate physical mechanism. The damage mecha-
nism is the result of collisions with atoms, which become displaced from the lattice,
creating interstitials and vacancies. Over time, sufficient displacement can occur and may change the performance properties of the device or material. Consequently, DDD is an effect of concern for all semiconductor bulk based devices. DDD accumulation primarily occurs when the semiconductor material is exposed to neutrons, trapped protons and solar protons over time. Likewise, secondary radiation produced in shielding materials can cause DDD effects.

As revealed by its name, the term “SEE” emphasizes the fact that the effect is caused by an individual particle interacting with the material. In current semiconductor technologies SEEs cause a much larger problem than the combination of all long-term cumulative effects. SEEs are explained in the next section.

**Single event effects**

SEEs are provoked by the strike of a single energetic particle (ion, proton, electron, neutron, etc.) in sensitive regions of the material. The particle travels through the semiconductor material, leaving an ionized track behind by depositing sufficient energy to generate an effect on a localized area of the electronic device. Although TID and SEE are both caused by ionizing radiation, their damage mechanisms are very different. While TID is a long-term effect that changes the electrical properties of the device, SEEs are the result of an instantaneous perturbation.

Neutron and alpha (α) particles are the most common sources of SEEs in terrestrial environments, while in space SEEs are primarily caused by cosmic rays and heavy ions. SEEs affect many different types of electronic devices and technologies, causing data corruption, high current conditions and transient disturbances. Left untreated, harmful functional interruptions and catastrophic failures can easily occur.

Semiconductor devices experience SEEs in two major forms: in the form of hard errors with destructive effects and in the form of soft errors with nondestructive effects. While the former leads to permanent degradation or even destruction of the device, the latter leaves no permanent damage.

Soft errors are of temporal nature, implying that the physical functionality of the circuit is not affected, even though its temporal integrity is. Typical examples include undesired changes of logic value in sequential logic and undesired analogue pulses that temporarily change the output of combinational logic. Soft errors can be further divided into transient and static errors (Mavis and Eaton, 2002). When a soft error occurs, it can result in a detected recoverable error, detected unrecoverable error, or silent data corruption (Kadayif et al., 2010; Weaver et al., 2004). There are numerous types of SEE, which can be categorized depending on the type of degradation, recoverability, and technology susceptibility.

2.3.4 **Redundancy in dealing with faults**

To increase the reliability of safety-critical systems, effective methods and techniques have to be employed to prevent or reduce the appearance of faults that could
cause catastrophic system failures. For system design, faults can be dealt with by utilizing two different design strategies: fault avoidance and FT.

Since a failure is the result of an error spreading, and an error is the result of a fault, eliminating faults would improve reliability. Targeting the source of faults, fault avoidance techniques attempt to prevent faults from occurring in the first place. Therefore, fault avoidance strategies can be used in the design stage of electronic devices. Examples of fault avoidance techniques are silicon on insulator and hardened memory cells. However, complete removal of faults via fault avoidance is not possible. Besides, these techniques and supporting technologies have drawbacks regarding cost, speed of operation, chip area, and power consumption.

Once a fault has been generated, it can be prevented from activating an error by using either static fault-tolerant techniques or dynamic fault-tolerant techniques. Hence, FT strategies are implemented at the time of job execution. As faults and tolerance to them can be considered at different abstraction levels of system description, FT strategies can be implemented at a system level or element level. Fig. 2.5 shows the failure life cycle and different techniques employed to deal with faults at each stage of development.

**Fault avoidance**

In mainstream systems, fault avoidance design strategies are employed to achieve the projected failure rates of the system. Manufacturing companies assess and identify the sources and weaknesses that could lead to potential failures. Based on the assessments, preventive measures are taken to ensure that the overall reliability target is met.

In addition, fault avoidance strategies may also lead to modifications of conventional manufacturing processes by using technology and design mitigation techniques. These techniques involve the use of specific materials, modification of the doping profiles of devices and substrates, and optimization of deposition processes for insulators.

More specifically, technology mitigation techniques are based at the process level, and consist of IC process variations by either improving the manufacturing process or by improving the materials used. Design mitigation techniques operate at
the layout level, such as the use of enclosed layout transistors. Furthermore, to pre-
vent or reduce the effects of radiation, memory cells can be hardened by using con-
tact and guard rings, and safety margins can be inserted into clock speed, operating
temperature and supply voltage margins.

**Fault tolerance by using redundancy**

The key strategy of FT is redundancy, which is defined as the addition of informa-
tion, resources or time beyond what is needed for correct system operation. Redundancy may include combinations of additional elements of HW and/or soft-
ware, and fault-tolerant techniques rely on them to detect and/or recover from
faults. These components are called redundancy because in a perfect system they
are not needed.

As a design strategy, artificially built-in or protective redundancy is considered
as a system property, which is the incorporation of extra components in the design
of a system so that its function is not impaired in the event of a failure. Redundancy
may arise by design (artificially built-in redundancy) or as a natural
by-product of design (natural redundancy). While the former has been deliberately
introduced, the latter is not usually exploited. Here, only artificial design redun-
dancy is considered.

When the minimum reliability of a system is not satisfied, extra redundancy, not
strictly necessary for the normal functioning of the system, can be added in order to
increase the probability of normal functioning. Note that redundancy indicates the
performance of the same task, not identical functionality. Therefore, heterogeneous
HW performing the same work can also provide redundancy.

FT is composed of actions such as fault detection, location of the faulty compo-
nent, recovery and, if necessary, reconfiguration of the system. Fault detection and
location mean, respectively, to determine the presence of faults and the time of
occurrence, and to exactly locate the reason/origin of the fault. It should be empha-
sized that system recovery must be performed dynamically during job execution as
if the system is “as good as new” in operational terms, even though some of the
redundancy has been used up and this may limit the possibilities for future repairs.

In classifying redundancy, this work follows the approach proposed by Schagaev
(2001). Fig. 2.6 shows the different types of redundancy and how it is implemented
in HW and software systems. In general, there are three types of redundancy: struc-
tural (S), involving multiplication of components, information (I), involving multipli-
cation of information/data and time redundancy (T), involving multiplication of
functions in time. Here, the focus is on the HW aspect of redundancy and FT.

**Structural hardware redundancy**

Structural HW redundancy involves multiple independent HW components and
assumes execution of the same computation over such components at the same
time. When this type of redundancy is used for reliability purposes the errors are
exposed by checking and comparing the results of the independent executions.
general, redundancy can be applied at different levels of abstraction: from transistor level, gate or logic level up to microcode level and chip level.

Two different organizations of structural redundancy can be distinguished:

- Parallel redundancy with redundant components running concurrently
- Standby redundancy with a spare component being activated upon failure of an active component

Structural redundancy can be divided into two types: static and dynamic.

1. Static redundancy

   Static redundancy, also called masking redundancy, performs error mitigation. The term “static” is linked to the fact that redundancy is built into the system structure. Fault-tolerant techniques based on this type of redundancy transparently delete errors on detection. Triple modular redundancy (TMR) (Von Neumann, 1956) and its generalization N-modular redundancy (NMR) represent the most common form of HW redundancy. Fig. 2.7 illustrates a TMR with a majority voter.

2. Dynamic redundancy

   This type of redundancy is similar to static redundancy, but its voter logic is replaced with a switch that is controlled by an error detection block. Dynamic redundancy techniques are used in order to reduce the extensive space, energy and performance overheads of TMR and NMR systems.
3. Hybrid redundancy

This type of redundancy approach is formed by mixing fault masking, detection, location, and recovery to combine the advantages of static and dynamic redundancy (Johnson, 1989). Hybrid approaches employ fault masking to prevent erroneous results from being processed and erroneous data being spread across the system. Hybrid techniques also use fault detection, location and recovery to improve FT by removing errors.

Information redundancy

Information redundancy denotes the addition of new information to existing information, i.e., using more bits than needed to ensure fault-free functioning. The most common form of information redundancy is coding. Coding consists of adding check bits to the data allowing (1) the verification of data correctness and/or (2) the correction of erroneous data. Coding theory was initially motivated by the need to mitigate errors in information transmission (Shannon, 1948) and thus has a long history of application.

Time redundancy

The main problem with the structural and information redundancy types discussed above is the penalty imposed in the form of extra HW. At the expense of using additional time, FT techniques based on time redundancy aim to reduce the amount of HW required for the implementation. Time redundancy techniques involve multiple executions of a program using the same piece of HW and comparing the execution results to determine if a fault has occurred. This approach is known to be effective in detecting errors caused by transient faults. There are various time redundancy techniques, such as using alternating logic, recomputing with shifted, rotated or swapped operands (the address in a computer instruction of data), etc.

2.3.5 A fault-tolerant system with enhanced resilience against disturbance

Assume $M$ as the known model of a system that performs a given function $F$. To analyses methods for achieving a required level of reliability, with performance and power consumption constraints, the following three models are defined:

- $M_s$: The model of the system
- $M_{\text{fault}}$: The model of the faults that a real-time FT system will be exposed to
- $M_{\text{FT}}$: The model of FT or the new structure that implements FT

As shown in Fig. 2.8, these models are mutually dependent. Note that in this approach solution costs are not considered. $M_{\text{fault}}$ represents all faults that a system must tolerate. Typical examples of HW faults include:

- Byzantine fault: the behavior of a component that gives conflicting values to other components; this fault affects the entire system.
Subsystem fault: a temporary or permanent incorrect behavior of a subsystem; this fault affects the entire system.

Stuck-at fault: the result value is fixed at 0 or 1.

Fail-stop fault: a faulty unit stops functioning, producing no bad output.

Hidden fault or latent fault: a behavior fault that exists in the HW over a long period of time, e.g., Byzantine fault and fail-stop fault.

Fault encapsulation approaches can facilitate fault handling: due to deliberate design solutions it is possible to ensure that severe faults in the system do not escalate and remain simpler to handle, therefore making the fault handling possible to implement in practice.

*Fig. 2.9* illustrates the concept of the new approach with a feature of high reliability that deals with various faults in the system and presents a variety of possible solutions. The system model in *Fig. 2.9* has overlapped HW and system software (SSW) ellipses to represent the duality of the system: HW and SSW. Both of them must be involved in the implementation of FT, as the HW cannot cover all possible faults. Obviously, $M_{FT}$ plays the role of a “conceptual deliverer” of reliability for the real-time FT system. Therefore, it has to be effective during the whole operational lifetime of the computer itself.

Contrary to the usual conception in reliability modeling, one has to assume that a fault might exist in the system over an arbitrary long period of time (latent fault) and its detection and elimination is not possible “at once.” Following Dijkstra’s approach (Dijkstra, 1965) of defining a function as a process described by its algorithm, here FT is also considered as a function described and implemented by an algorithm.

---

**Figure 2.8** New feature of an FT system for achieving high reliability.

**Figure 2.9** A model of a computer system with the new FT feature.
There are several options to achieve FT, assuming the use of HW and SSW by employing various types of redundancy techniques mentioned above. However, the use of a certain type of redundancy might cause system performance degradation; this is especially true for software measures (Kulkarni et al., 1987; Oh et al., 2002).

**Fault model**

It is not realistic to describe all possible faults that may occur in a system individually. Hence, fault models (FMs) are developed and used to represent in a simple form the consequence of complex physical mechanisms that lead to faults. Faults are assumed to behave according to these FMs in order to make their evaluations possible. According to Dunn (1991), a FM is considered as a way of summarizing many fault descriptions at once. Often, it is desirable to model many different faults simultaneously by summarizing their common characteristics.

In the case of electronic systems, modeling of faults can be implemented at two different levels: at the level of HW components or at the system level. The following classification of faults is primarily based on the work by Avizienis et al. (2004).

1. Classification of faults by origin
   - Faults can be classified based on the attributes related to their origin, including their cause, the level at which they occur, the phase of creation, nature, system boundaries, etc.

2. Classification of faults by manifestation
   - Faults can also be classified based on the attributes related to their manifestation, including their response, dimension, reproducibility, extent, persistence, etc.

**Fault tolerance and system modeling**

When a fault occurs, extra redundancy is required to deal with it. Thus, redundancy and the ability to use it to remove the fault form a combination of tools and techniques that is required for implementing reconfiguration.

When system and fault modeling are developed together, system behavior in the presence of faults and the control process of FT can be contemplated at earlier stages, taking into account mutual dependencies of solutions at every stage of the design and development process. This advantage can be fully exploited in designing embedded systems, which have many design constraints and very often mutually exclusive requirements in reliability, performance, and power consumption that can limit design options.

Note that redundancy can be used for various purposes, and it can be involved as an essential part of reconfiguration of connected computers. System reconfiguration purposes include performance improvement, reliability enforcement and energy-wise use.

**Generalized algorithm of fault tolerance**

Although FT in computer systems can be achieved by introducing static redundancy in HW and SSW, employing these traditional approaches is expensive in terms of
time, information or HW overheads. To tackle these problems, Schagaev and Sogomonian (Schagaev, 1986a,b, 1987; Sogomonian and Schagaev, 1988) proposed to consider FT not only as a feature, but also as a process that can be implemented algorithmically. Fig. 2.10 shows a process for implementing FT, which assumes dynamic interaction of existing redundancy types between elements. The original three-step algorithm (Sogomonian and Schagaev, 1988) has been further developed into the generalized algorithm of FT (GAFT), a five-step fault-handling algorithm shown in Fig. 2.10:

Detecting faults
Identifying faults
Identifying faulty components
Hardware reconfiguration to achieve a repairable state
Recovery of a correct state(s) for the system and user software.

The different types of redundancy (information, time, and structural either HW- or software-based) can be used to implement every step of the GAFT. As transient faults occur an order of magnitude more often than permanent faults, handling of transient FT must be extremely effective. A good fault-tolerant system tolerates the vast majority of transient faults within the interval of instruction execution, making them invisible for other instructions. At the same time, when transient faults with longer time range or permanent faults are encountered, they might be detected and recovered differently, at the procedural or task level of SSW.

It should be emphasized that the dynamic interaction feature of the GAFT is vital to building a fault-tolerant system, which is well illustrated in Fig. 2.11. Fig. 2.11 compares the traditional approach with the GAFT approach in dealing

---

**Figure 2.10** A dynamic process for implementing FT using the Generalized Algorithm of Fault Tolerance (GAFT).

*Source:* Adapted from Fig. 5.11 in Castano, V. and Schagaev, I. (2015). Resilient Computer System Design. Springer.
with faults, where $S$ represents one of the five possible states for a single processing element system: ideal, faulty, erroneous, degraded and failed; $T$ shows transitional operations and $M$ denotes mechanisms involved in FT. While in the GAFT approach FT is carried out in a dynamic process involving interactions between different elements, in the traditional approach a fault is treated as a single event and error mitigation is performed statically. Obviously, the GAFT strategy represents a resilient approach to handling faults and disturbances.

### 2.3.6 Hardware and system software support of resilience

The development of this new computer architecture follows the holistic principles proposed by Schagaev et al. (2010): simplicity, redundancy, reconfigurability,
scalability, reliability, and resource awareness. A brief introduction to these principles is given below:

- **Simplicity**: Complexity is difficult to implement and handle efficiently. In addition, big complex systems are more prone to faults, thus lowering reliability.
- **Reliability**: The highest reliability of individual components is preferable, but always keeping in mind the cost efficiency of its implementation.
- **Redundancy**: Deliberate introduction of HW and software redundancy provides the required level of reconfigurability to reach performance and reliability goals.
- **Reconfigurability**: Apart from the simplicity, reliability and deliberate introduction of redundancy, it is essential to achieve balance between performance, reliability, and power. Reconfigurability serves three main purposes: performance, reliability, and power awareness. It allows the system to adapt twofold: first to recover from a permanent fault and second adjusting the requirements of the running application.
- **Scalability**: Scalability should be kept in mind when designing a system so that it can be extended if the requirements change.
- **Power-awareness**: Mission critical systems have significant limitations of HW resources and power consumption constraints (e.g., battery life). Thus, for wise resource use, reconfigurability must be introduced.

By following these principles a new HW architecture and SSW have been developed for safety-critical applications, a resilience approach to change and disturbance with the ultimate goal of delivering correct service within specified time constraints. Further discussion on the HW and SSW development is deemed out of scope.

**References**

Adams, J.H., Silberberg, R., Tsao, C.H., 1982. Cosmic ray effects on microelectronics. IEEE Trans. Nucl. Sci. 29, 169—172.

Antola, A., Erényi, I., Scarabottolo, N., 1986. Transient fault management in systems based on the AMD 2900 microprocessors. Microprocess. Microprogram. 17, 205—217.

Asanovic, K., Bodik, R., Catanzaro, B., et al., 2006. The Landscape of Parallel Computing Research: A View from Berckene. Technical Report No. UCB/EECS-2006-183.

Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C., 2004. Basic concepts and taxonomy of dependable and secure computing. IEEE Trans. Dependable Secure Comput. 1, 11—33.

Barth J.L., LaBel, K.A., Poivey, C., 2004. Radiation assurance for the space environment. In: Integrated Circuit Design and Technology, ICICDT’04, pp. 323—333. doi: 10.1109/ICICDT.2004.1309976.

Baumann, R.C., 2002. The impact of technology scaling on soft error rate performance and limits to the efficacy of error correction. In: International Electron Devices Meeting (IEDM’02), Digest., pp. 329—332.

Baumann, R.C., 2005a. Soft errors in advanced computer systems. IEEE Des. Test Comput. 2, 258—266.

Baumann, R.C., 2005b. Radiation-induced soft errors in advanced semiconductor technologies. IEEE Trans. Device Mater. Reliab. 5, 305—316.

Binder, D., Smith, E.C., Holman, A.B., 1975. Satellite anomalies from galactic cosmic rays. IEEE Trans. Nucl. Sci. 22, 2675—2680.
Blake, J.B., Mandel, R., 1986. On-orbit observations single event upset in Harris HM-6508 1K RAMS. IEEE Trans. Nucl. Sci. 33, 1616–1619.
Breuer, M.A., 1973. Testing for intermittent faults in digital circuits. IEEE Trans. Comput. C. 22, 241–246.
Castano, V., Schagaev, I., 2015. Resilient Computer System Design. Springer.
Constantinescu, C., 2003. Trends and challenges in VLSI circuit reliability. IEEE Micro. 23, 14–19.
Dijkstra, E.W., 1965. Solution of a problem in concurrent programming control. Commun. ACM. 8, 569.
Dunn, M., 1991. Designer fault models for VLSI. IEE Colloquium Des. Testability (Digest No. 1991/102). 1991, 4/1–4/5.
Felix, J.A., Shaneyfelt, M.R., Schwank, J.R., Dalton, S.M., et al., 2007. Enhanced degradation in power MOSFET devices due to heavy ion irradiation. IEEE Trans. Nucl. Sci. 54, 2181–2189.
Hazucha, P., Svensson, C., 2000. Impact of CMOS technology scaling on the atmospheric neutron soft error rate. IEEE Trans. Nucl. Sci. 47, 2586–2594.
Hazucha, P., Karnik, T., Maiz, J., Walstra, S. et al., 2003. Neutron soft error rate measurements in a 90-nm CMOS process and scaling trends in SRAM from 0.25-µm to 90-nm generation. In: Technical Digest, IEEE International Electron Devices Meeting (IEDM’03), pp. 21.5.1–21.5.4.
Johnson, B.W., 1989. The Design and Analysis of Fault Tolerant Digital Systems. Addison-Wesley, Reading, MA.
Kadayif, I., Sen, H., Koyuncu, S., 2010. Modelling soft errors for data caches and alleviating their effects on data reliability. Microprocess. Microsyst. 34, 200–214.
Karnik, T., Hazucha, P., 2004. Characterization of soft errors caused by single event upsets in CMOS processes. IEEE Trans. dependable Secure Comput. 1, 128–143.
Kish, L.B., 2002. End of Moore’s Law: thermal (noise) death of integration in micro and nano electronics. Phys. Lett. A. 305, 144–149.
Kulkarni, G.V., Nicola, F.V., Trivedi, S.K., 1987. Effects of Checkpointing and Queuing on Program Performance. Duke University, Durham, NC.
Mavis, D.G. and Eaton, P.H., 2002. Soft error rate mitigation techniques for modern microcircuits. In: 40th Annual Reliability Physics Symposium Proceedings, pp. 216–225.
Miller, L.A., Brice, D.K., Prinja, A.K., Picraux, S.T., 1994. Molecular dynamics simulations of bulk displacement threshold energies in Si. Radiat. Eff. Defects Solids: Incorporat. Plasma Sci. Plasma Technol. 129, 127.
Moore, G.E., 1965. Cramming more components onto integrated circuits. Electronics. 38 (8), 114–117.
Oh, N., Shirvani, P.P., McCluskey, E.J., 2002. Control-flow checking by software signatures. IEEE Trans. Reliab. 51, 111–122.
Schagaev, J.Z.I., 2001. Redundancy classification and its applications for fault tolerant computer design. IEEE TESADI-01, Arizona, Tucson.
Schagaev, I., 1986a. Detecting the defective computer in two-unit, fault tolerant system having a sliding stand-by units. Autom. Remote Control. 5, 143–150.
Schagaev, I., 1986b. Algorithms of computation recovery. Automatic and Remote Control 7. Plenum, New York, NY.
Schagaev, I., 1987. Algorithms to restoring a computing process. Automatic and Remote Control. Plenum, New York, NY, p. 7.
Schagaev, I., Kaegi, T., Gutknetch, J., 2010. ERA: evolving reconfigurable architecture. In: Proceedings of 11th ACIS, Presented at the International Conference on Software.
2.4 Mathematical generalizations of resilience theories and a two-step solution of nonlinear dynamical behavior

This section first presents a mathematical generalization of the resilience theories developed in various fields of study. Then, a theoretical work for solving nonlinear dynamic problems using a generalized two-step approach is reviewed for possible applications in complex dynamical systems.

2.4.1 A mathematical definition of resilience

Perhaps the fundamental issue of the resilience approach is best revealed by McDonald’s answers to his own question: What is the nature of resilience? (McDonald, 2006):

If resilience is a system property, then it probably needs to be seen as an aspect of the relationship between a particular socio-technical system and the environment.

References

- Seifert, N., Zhu, X., Massengill, L.W., 2002. Impact of scaling on soft-error rates in commercial microprocessors. IEEE Trans. Nucl. Sci. 49, 3100–3106.
- Shannon, C.E., 1948. A mathematical theory of communication. Bell. Syst. Tech. J. 27 (379–423), 623–656.
- Shivakumar, P., Kistler, M., Keckler, S.W., Burger, D., Alvisi, L., 2002. Modeling the effect of technology trends on the soft error rate of combinational logic. In: Proceedings of the International Conference on Dependable Systems and Networks (DSN 2002), pp. 389–398.
- Sogomonian, E., Schagaev, I., 1988. Hardware and software fault tolerance of computer systems. Avtomatika i Telemekhanika, 3–39.
- Von Neumann, J., 1956. Probabilistic logics and synthesis of reliable organisms from unreliable components. In: Shannon, C., McCarthy, J. (Eds.), Automata Studies. Princeton University Press, pp. 43–98.
- Weaver, C., Emer, J., Mukherjee, S.S., Reinhardt, S.K., 2004. Techniques to reduce the soft error rate of a high-performance microprocessor. In: Proceedings of the 31th Annual International Symposium on Computer Architecture, pp. 264–275.
- Yu, Q.K., Tang, M., Zhu, H.J., Zhang, H.M. et al., 2005. Experimental investigation of radiation damage on CCD with protons and cobalt 60 gamma rays. In: 8th European Conference on Radiation and Its Effects on Components and Systems (RADECS 2005), pp. LNW3-1–LNW2-5.
of that system... Resilience represents the capacity (of an organisational system) to anticipate and manage risk effectively, through appropriate adaptation of its actions, systems and processes, so as to ensure that its core functions are carried out in a stable and effective relationship with the environment.

This in-depth analysis summarizes all the fundamental aspects of the resilience theories reviewed above, clarifying the ultimate goal of resilience studies as to ensure a stable and effective relationship between the system and the system environment. His description contains four key ingredients: a system, the system environment, the relationship between the two, and a risk or disturbance. A mathematical generalization of resilience is given below.

A system is defined as $S(q_1, q_2, \ldots, q_n)$, and $[q]$ are key system variables. The system environment is expressed as $V$, which is directly influenced by the key system variables. A disturbance is denoted by $D$. An impact function is defined as $I(q_1, q_2, \ldots, q_n)$, which is a function of the key system variables and measures the impact of a disturbance to the system on the system environment. This impact function represents the relationship between the system and the system environment, and is assumed to be a smooth function.

When a disturbance occurs, it releases a certain amount of destructive energy, $E_d$, into the system. Let $E_s(1), E_s(2)$, represent, respectively, the energy absorbed by the system and the inner energy released by the system into the system environment during or after the event.

Then, the total energy absorbed by the system environment, $E_v$, is obtained as

$$E_v = E_d - E_s(1) + E_s(2) \quad (2.1)$$

When facing a disturbance, a resilient system completely absorbs the energy released by an adverse event, and prevents its potentially harmful system energy from being released into the system environment (e.g., the nuclear meltdown of a nuclear power plant, the toxic chemical pollution of a chemical plant, the crash of a commercial airplane), i.e.,

$$E_d = E_s(1) \quad (2.2)$$

$$E_s(2) = 0 \quad (2.3)$$

Substituting Eqs. (2.2) and (2.3) into Eq. (2.1) leads to

$$E_v = 0 \quad (2.4)$$

With no harmful energy being released into the system environment, the impact function retains its minimum value, i.e.,

$$\frac{dI(q)}{dq} \bigg|_{[q]=[q^*]} = 0 \text{ or } I'(q^*) = 0 \quad (2.5)$$
Here, the characteristic system variables \([q^*]\) represent the resilience design of a system. Fig. 2.12 shows the concept of the impact function of the system environment and examples of resilient and nonresilient systems. In general, a system design is subjected to various constraints or restrictions, and thus even a resilient system may exert a limited impact on the system environment, which is represented by \(I_0\) in Fig. 2.12. Nevertheless, compared with any other approaches that do not incorporate resilience strategies, the impact of a resilient system on the system environment is at the minimum.

Obviously, for a socioecological system or a sociotechnical system, the system environment of that local system refers to the much larger socioecological or sociotechnical environment interacting with that system, including the people who live and work in it. For an embedded computer system used in a pacemaker implanted in a heart-disease patient, the patient is the system environment; if it is used in a commercial airplane, the passengers and the crew are the immediate system environment, and so forth. The key system variables are the controlling system variables, through which the seven resilience principles of the socioecological system, the four essential resilience capabilities of the sociotechnical system, and the six resilience attributes of the computer system can be introduced into the system.

Finally, a mathematical definition of resilience is attempted. In the event of any disturbance, if a system completely absorbs the destructive energy released by that event, i.e., \(E_d = E_s(1)\), and does not release any harmful system energy into the system environment as a result of that event, i.e., \(E_s(2) = 0\) and \(E_v = 0\), then the impact function of the system environment retains its minimum value, satisfying \(I'(q^*) = 0\). Under this condition of minimum impact, a system is defined as a resilient system.

Based on this definition, the priority of the resilience approach is to minimize the negative impact of an adverse event occurring to a system on the system environment. Although there is no explicit mention of the system condition during or

![Figure 2.12](image-url)  
**Figure 2.12** Concept of the impact function of the system environment and resilient and nonresilient systems subjected to adverse events.
after that event, it is implicitly defined by the system’s energy conditions specified in Eqs. (2.2) and (2.3). Obviously, these conditions can be met only when a system possesses the resilience characteristics specified by the various resilience theories discussed so far.

The impact-function-based definition of resilience focuses on the various energies exchanged between a disturbance and a system, and between the system and the system environment during an adverse event. This definition enables a unified physical interpretation of resilience in various fields of study. Several general examples of resilient and nonresilient systems with energy conditions specified are given below:

1. In earthquake engineering, a resilient community is one that sustains limited damage despite a major earthquake, and is able to recover its normal function quickly, thus satisfying the conditions $E_d = E_s(1)$, $E_s(2) = 0$ and $E_v = 0$.

2. A resilient computer system for safety-critical applications must possess the resilience attributes described in Section 2.3. Such a system satisfies the conditions $E_d = E_s(1)$, $E_s(2) = 0$ and $E_v = 0$ in the event of faults and errors, such as those caused by radiation, by using any one of the fault-tolerant and error-mitigation techniques discussed previously.

3. The two nuclear accidents, one at the Chernobyl nuclear power plant in 1986 and the other at the Fukushima nuclear power plant in 2011, proved that the two power plants were nonresilient systems. Even though the causes of these accidents were completely different [the former due to a flawed reactor design with operation by inadequately trained personnel (WNO, 2016), and the latter due to a natural disaster], the irreversible, worst-case conditions of $E_s(2) > 0$ and $E_v > 0$ (or $E_s(2) = \infty$ and $E_v = \infty$) occurred in both cases.

The impact function will be discussed further in the next chapter. Defining the ratio of $E_s(1)$ to $E_d$ as a resilience index, the impact function will be redefined as a function of this resilience variable. The specific form of the impact function will also be discussed.

### 2.4.2 A generalized two-step solution of nonlinear dynamical behavior

In principle, exact solutions of complex dynamical systems that are the targets of resilience studies cannot be found, with respect to the equations of motion of systems that are often mechanical or physical in nature, as well as systems that arise in biology, economics, and elsewhere. This is because a precise mathematical description of the system behavior will inevitably lead to ill-defined dynamic equations that cannot be solved. Nevertheless, knowledge of the long-term qualitative behavior of a complex dynamical system is highly desirable for developing an efficient and reliable resilience approach.

In the following review, a generalized two-step approximate solution to a nonlinear dynamic problem (Shi et al., 2014) is introduced first, aiming at its wider application to similar types of problems in complex dynamical systems. Then, this
method is applied to studying the cracking behavior of a large concrete dam during a strong earthquake, and the resilience issue of a complex dam-reservoir system against natural disasters is briefly discussed.

**Generalized two-step solution**

To generalize the solution strategy of the two-step approach, a dynamic response problem in association with a time-dependent excitation or energy input is defined by using four functions, which are the target function, \( T(x, y, z) \); the excitation or energy function, \( E(t) \); the force function of \( T(x, y, z) \), \( F_T(x, y, z, t) \); and the deformation function of \( T(x, y, z) \), \( D_T(x, y, z, t) \). In general, the purpose of a linear or nonlinear response analysis of \( T(x, y, z) \) subjected to \( E(t) \) is to obtain \( D_T(x, y, z, t) \).

The target function \( T(x, y, z) \) is a well-defined space function, representing a target of study such as a dam. The excitation function \( E(t) \) stands for a source of time-dependent energy input, such as an earthquake, and subjecting \( E(t) \) to \( T(x, y, z) \) generates a dynamic response in the target function. The time-varying dynamic force function \( F_T(x, y, z, t) \) is derived from the dynamic response of \( T(x, y, z) \), such as the function of inertia force in a dam under earthquake force. Acted upon by this force function, a deformation or variation of \( T(x, y, z) \) occurs, which is expressed by the deformation function \( D_T(x, y, z, t) \). The aim of the response analysis is to obtain this time-dependent deformation function, from which important features in the dynamic response of the target function are studied, such as the cracking behavior of a dam during a large earthquake.

In mathematical terms, the above description of the problem may be expressed as

\[
T(x, y, z) \times E(t) \rightarrow F_T(x, y, z, t) \times D_T(x, y, z, t). \tag{2.6a}
\]

or

\[
T(x, y, z) \times E(t) = F_T(x, y, z, t) \times D_T(x, y, z, t). \tag{2.6b}
\]

For linear response:

\[
T(x, y, z) \times E(t) = F_T(x, y, z, t)|_{L} \times D_T(x, y, z, t)|_{L} \tag{2.7}
\]

In general, an accurate solution of Eq. (2.7) is possible.

For nonlinear response:

\[
T(x, y, z) \times E(t) = F_T(x, y, z, t)|_{NL} \times D_T(x, y, z, t)|_{NL}. \tag{2.8}
\]

In general, an accurate solution of Eq. (2.8) is impossible. This is because the nonlinear deformation function \( D_T(x, y, z, t)|_{NL} \) contains elements of change that can fundamentally transform the target function \( T(x, y, z) \) from its original form,
such as a gradual change in the geometry or mass, or some other basic features, resulting in ill-defined dynamic equations that cannot be solved.

The two-step approach suggests obtaining an approximate function $D^*_T(x, y, z, t)|_{NL}$ for $D_T(x, y, z, t)|_{NL}$ based on $F_T(x, y, z, t)|_L$, assuming that $F_T(x, y, z, t)|_L$ is a close approximation to $F_T(x, y, z, t)|_{NL}$, i.e.,

$$T(x, y, z) \times E(t) \cong F_T(x, y, z, t)|_L \times D^*_T(x, y, z, t)|_{NL}. \quad (2.9)$$

Eq. (2.9) transforms the time history analysis in Eq. (2.8) from dynamic analysis to static analysis based on the known function of $F_T(x, y, z, t)|_L$, and it is believed that this redefined nonlinear problem can be accurately solved to obtain the function $D^*_T(x, y, z, t)|_{NL}$. The logical reasoning based on mechanics principles for this approach is that, if the linear force function $F_T(x, y, z, t)|_L$ is a close approximation to the actual force function $F_T(x, y, z, t)|_{NL}$, then the approximate deformation function $D^*_T(x, y, z, t)|_{NL}$ must be a good estimate of the actual deformation function $D_T(x, y, z, t)|_{NL}$. Fig. 2.13 illustrates the flow of this two-step solution. The error of the approach may be estimated in percentage form as

$$\text{Error} = \left| \frac{W^*_F - W_F}{W_F} \right| \times 100\% \quad (2.10)$$

![Figure 2.13 Concept of the two-step approach to nonlinear dynamic problems.](image)
Here, \( W_F \) is the total work spent in the process (such as the total fracture energy consumed for propagating all the cracks in a dam during an earthquake) that gives rise to the important features in \( D_T(x, y, z, t)|_{NL} \) which are the focus of the study, and \( W_F^* \) is the corresponding value calculated from \( D^*(x, y, z, t)|_{NL} \).

For certain types of nonlinear response phenomena, due to some specific feature of the problem in question the nonlinearity in its deformation function \( D_T(x, y, z, t)|_{NL} \) may have little influence on the system’s dynamic response (not vice versa; after all, the nonlinearity is caused by the dynamic response). Hence, the force function obtained by linear response analysis, \( F_T(x, y, z, t)|_L \), is likely to represent closely the actual force function, \( F_T(x, y, z, t)|_{NL} \), which in principle cannot be accurately derived from nonlinear response analysis.

Finally, it is worth pointing out that though the target function \( T(x, y, z) \) is initially assumed to be independent of time; actually, it can also contain a time variable such as \( T(x, y, z, t) \), provided that this new target function is also a known function in terms of space and time.

**Resilience assessment of a complex dam-reservoir system**

Figs. 2.14 and 2.15 show successful applications of this two-step solution in obtaining the discrete cracking behavior of a large concrete gravity dam based on the linearly obtained time history of the dam’s inertia force field under inland type and oceanic-trench type earthquake motions. Note that in order to propagate large cracks in the dam two extremely strong earthquakes are assumed in the study. With

![Figure 2.14](image_url) Display of continuously changing crack behavior within a split second during the rescaled Hitokura earthquake (inland type) (Shi et al., 2014).
reference to a detailed study on the seismic cracking behavior of concrete dams (Shi et al., 2014), the focus of discussion here is on the resilience issue of a dam-reservoir system, using the newly proposed resilience theory discussed in the preceding section.

Based on the obtained cracking behavior, the resilience of the dam-reservoir system against disturbance can be classified into three categories:

1. A resilient system:
   Despite a large earthquake, the dam has sustained no serious cracking damage that could affect the normal functioning of the dam-reservoir system. Thus, the conditions \( E_d = E_{s(1)} = 0 \) and \( E_v = 0 \) are satisfied, i.e., there is no harmful energy released into the system environment. Here, \( E_d \) and \( E_{s(1)} \) represent, respectively, the destructive earthquake energy and the kinematic energy of the dam during seismic motions.

2. A nonresilient system:
   As a result of a large earthquake, the dam has sustained severe cracking damage that could affect the normal functioning of the dam-reservoir system, such as flood control due to damage to the spillway. Note that seismically damaged dams can require several years before complete repair and functional restoration can be accomplished (Hartford, 2011). In this case, \( E_{s(1)} < E_d \), \( E_{s(2)} = 0 \) and \( E_v > 0 \), implying a harmful effect on the system environment. Obviously, the fracture energy sustained by the dam under seismic forces, \( W_F \), represents that harmful effect, i.e., \( W_F = E_d - E_{s(1)} \) and \( E_v = W_F \).

3. A failed system:
   As a result of a large earthquake, the dam has failed catastrophically due to the occurrence of penetrating cracks, causing grave social, environmental and economic consequences to the downstream areas. In this case, obviously, \( E_{s(2)} = \infty \) and \( E_v = \infty \).

Figure 2.15 Display of continuously changing crack behavior within a split second during the rescaled Urakawa earthquake (oceanic-trench type) (Shi et al., 2014).
References

Hartford, D.N.D., 2011. Combinations of earthquake and flood hazards together with other factors. In: Schleiss, A.J., Boes, R.M. (Eds.), Dams and Reservoirs under changing challenges. Taylor & Francis Group, London, pp. 685–692.

McDonald, N., 2006. Organisational resilience and industrial risk. In: Hollnagel, E., Woods, D.D., Leveson, N. (Eds.), Resilience Engineering: Concepts and Precepts. Ashgate, Aldershot, UK, pp. 155–180.

Shi, Z., Nakano, M., Nakamura, Y., Liu, C., 2014. Discrete crack analysis of concrete gravity dams based on the known inertia force field of linear response analysis. Eng. Fract. Mech. 115, 122–136.

WNO, 2016. <www.world-nuclear.org/information./chernobyl-accident.aspx>.