SURVEY

Ecological thresholds: a survey

Roldan Muradian

Department d'èconomia i d'Histo`ria Econo`mica, Universitat Autònoma de Barcelona, Edifici B, 08193 Bellaterra, Barcelona, Spain

Received 16 May 2000; received in revised form 14 December 2000; accepted 15 December 2000

Abstract

The existence of ecological discontinuities and thresholds has been recognised by ecological economics as a key feature to take into account in the study of environment–economy interactions. This paper reviews some theoretical developments and empirical studies dealing with ecological phenomena involving non-linear dynamics. The literature about this issue reveals that there is abundant evidence of discontinuities and threshold effects as the consequence of human perturbations on ecological systems. However, due to the complexities involved, the predictive capacity of ecology is limited and large uncertainties still remain. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Ecological discontinuities; Ecological thresholds; Alternative states; Ecosystem resilience; Stable equilibria

1. Introduction

Vis-a-vis neo-classical economics, ecological economics has paid more attention to feedback effects, self-organisation, uncertainties and non-linear dynamics. From a co-evolutionary and adaptive perspective (Norgaard, 1992), ecological discontinuities, being an important source of uncertainty, are key to understanding the complex interactions between the economy and the biophysical environment where it is embedded. The ecological economics literature about this topic has put emphasis on the formulation of environmental policy or management models as well as the proposition of adequate institutional responses when ecological thresholds are likely. However, there have not been many attempts to integrate the ecological literature into this discussion. This paper tries to review the theoretical developments and empirical studies dealing with discontinuities in the ecological science. It pretends to systematise useful ecological information for economists and policy-makers interested in non-linear economy–environment interactions. How often do ecological discontinuities occur in nature? How uncertain are they? Are the underlying mechanisms well understood? Around these kinds of questions some fruitful dealings between economists and ecologists may arise. The present paper pretends to collaborate with this dialogue.

Ecological discontinuities can be defined as a sudden change in any property of an ecological
system as a consequence of smooth and continuous change in an independent variable. Ecological discontinuities imply critical values of the independent variable around which the system flips from one stable state to another, that is, ecological thresholds. The definition of thresholds can be arbitrary in the sense that it depends on the temporal and spatial scale adopted. In general, no description in ecology makes sense without reference to particular temporal and spatial scales (Levin, 1992). Ecological discontinuities are phenomena of special interest for economics when they comprise abrupt changes in the ecological services provided by the involved ecological system (Myers, 1996; Daily, 1997; Pimentel et al., 1997). These services can be supplied by different levels of the hierarchical organisation of ecological systems. Therefore, the most suitable temporal and spatial scales to address ecological discontinuities can vary broadly depending not only on the ecological system involved but also on the dynamics of the environmental services provided by it.

For the current theory of environmental externalities, to take into account ecological discontinuities is not a serious problem. Although present models usually assume a continuous relationship between production and environmental externalities (Turner, 1999), it would be not complicated to relax this assumption and to assume the possibility of discontinuities in the external cost curve. Rather, ecological discontinuities are a challenge for the economic theory because the threshold values and the magnitude of the change are generally uncertain. Perrings and Pearce (1994) deal with this problem by proposing a penalty to be paid when an environmental standard is surpassed. This standard would be a value of the independent variable that imposes a limit to the economic activity in order to maintain the risk of surpassing the ecological threshold at acceptable levels. The aim of this measure is to assure the resilience of ecosystems, which allows statistically predictable environmental externalities. This notion is based on the belief that once the ecological threshold is surpassed, the general equilibrium effects cannot be estimated ex ante. This kind of approach, based on safe-minimum standards, seems more suitable than market mechanisms to deal with complex economy–environment interactions because prices are unable to detect when a system is approaching a threshold (Bishop, 1993; Folke, 1999). However, if the critical values which may trigger an ecological discontinuity are basically uncertain, the safe-minimum standards can be very elusive and the process of standard-definition very conflictive. Hence, ecological research on the occurrence and predictability of ecological discontinuities can be a crucial input to the environmental decision-making process. The ‘quality’ of this input will determine the relevance of scientific advisory in the decisional process (Functowicz and Ravetz, 1994). Because of that, it is helpful to check the ‘state of the art’ of the ecological science on this subject.

The existence of ecological discontinuities is also a key feature to take into consideration for designing environmental management regimes. Perrings and Walker (1997) have shown, at a theoretical level, how different management strategies may influence non-linear and long-term dynamics of ecosystems. They point out that optimal management should be sensitive to key variables conditioning long-term resilience of the system, like fire frequency in grasslands (see below). Gunderson et al. (1997) introduce a set of propositions about the adequate scales of analysis, institutional strategies and variable characterisation and monitoring to deal with uncertainty linked to non-linear behaviour in natural systems. The aim of the present paper is not to go further on these issues, but instead to visualise the current state of knowledge about ecological discontinuities.

The second and third sections of this paper link the concept of ecological thresholds to the debate on the relationship between biodiversity and (a) resilience and stability of ecosystems and (b) ecosystem functions, respectively. Then, a fourth section reviews some examples of multiple states and ecological discontinuities in each of the current most important human impacts on ecosystems, population harvesting, pollution, habitat fragmentation, ecosystem management and biological invasions. The paper finishes with some concluding remarks.
2. Biodiversity, resilience and stability

The reference to the maintenance of ecosystems’ resilience as a key aspect of sustainability and as a crucial aim for environmental policy has been very present in the ecological economics literature (Common and Perrings, 1992; Arrow et al., 1995; Folke et al., 1996; Ludwig et al., 1996; Berkes et al., 1998; Levin et al., 1998). According to this position, ecological discontinuities may arise when human intervention has reduced enough of the ecosystem’s resilience. When the system loses resilience it becomes vulnerable to perturbations that earlier could be absorbed without structural change (Gunderson et al., 1997).

Resilience has been defined in very different ways, but mainly two connotations dominate the ecological literature. The most common definition is related to the capacity of ecological systems to recover from a disturbance (Walker, 1995). Under this point of view, resilience can be measured by how fast the variables of the system return to their equilibrium following a perturbation (Pimm, 1984; MacGillivray et al., 1995). The second definition (the most used in the ecological economics literature) emphasises the existence of alternative states in ecosystems. Under this approach, resilience is the ability of the system to maintain its structure and patterns of behaviour in the face of disturbance, that is, the capacity to absorb perturbations and still persist (Holling, 1973, 1986). In this context, resilience can be measured by the magnitude of disturbance that can be absorbed before the system flips to another state — before an ecological threshold is reached. The former definition requires that at least one equilibrium or stable state exits and can be recognised. The second one assumes a set of variables through which persistence can be assessed.

In the ecological literature, resilience traditionally has been associated with the concept of ecological stability, one of the most nebulous terms in ecology. Grimm and Wissel (1997) have counted at least 163 definitions of stability concepts. However, despite this prolific production, only few basic components of the concept can be recognised. One is related to the permanence of a certain reference state, dynamics or a set of properties that identify the system. The other component is associated with the capacity to return to the reference state (or dynamics) after a disturbance. Thus, each of the previously mentioned concepts of resilience tackles complementary aspects of stability. Therefore, in order to avoid confusion, they should be renamed. Grimm and Wissel (1997) try to do this, proposing a new vocabulary in the ecological stability discussion. Under their classification, persistence and elasticity can be new names for the previous definitions of resilience.

Currently, the Earth is witnessing one of the highest rates of species extinction in its history (Ehrlich and Wilson, 1991; Myers, 1993; Smith et al., 1993; Hughes et al., 1997). Biological diversity decline has been proposed as a cause of ecosystem resilience loss (Walker, 1995). Starting from this proposition, for some authors, there exists a threshold of biodiversity below which ecosystems lose the self-organisation that enables them to provide ecological services (Perrings and Opschoor, 1994). From this viewpoint, it is crucial for environmental policy to understand the relationship between species richness and ecosystem stability. This theme has been a topic of interest for ecology since the 1950s. At that time, Elton (1958), based on some empirical evidence, naturalist observation and mathematical models, stated that complexity, measured as species richness, number of biological interactions (connectance) or interaction strength, begets stability. Moreover, MacArthur (1995) pointed out that the more possible energy pathways in a community, the less likely would be a change in species density as a consequence of other species outbreak or deletion. Elton’s and MacArthur’s propositions were adopted as ‘conventional wisdom’ until the 1970s, when theoretical models found that food web local stability1 was inversely correlated to the number of species in the community (May, 1973). However, mathematical models dealing with the return speed to initial equilibrium did not show a unique relation between this variable and a community’s complexity (Pimm, 1984).

---

1 A system is locally stable when it returns to the initial equilibrium following a small perturbation.
After the mid 1970s, research on the diversity–stability problem started to be empirical (rather than comparative or theoretical). Usually, the strategy of these studies has been to estimate the ability of different plant communities (varying in species richness) to maintain total plant biomass after a perturbation. In general, biomass stability is positively correlated to species richness, which apparently contradicts the theoretical results. This relationship has been found for different kinds of natural perturbations, highly variable rainfall (McNaughton, 1997); grazing (McNaughton, 1993) and severe droughts (Leps et al., 1982; Frank and McNaughton, 1991). A positive relationship between stability and diversity also has been found for other communities’ aggregated parameters. For example, McGrady-Steed et al. (1997) report that for nutrient uptake and community respiration, variability decreases as biodiversity increases. Nonetheless, not all empirical studies have found a positive relationship between species richness and ecosystem stability. Some works show that the positive relationship between biomass stability and diversity is not maintained at trophic levels higher than primary producers (Hurd and Mellinger, 1971). On the other hand, using artificial seeded plant communities and based on measurements of the soil microbial respiration:biomass ratio, Wardle and Nicholson (1996) concluded that stability does not respond predictably to shifts in species richness.

Tilman (1996) reports empirical results where species diversity was positively correlated to stability of total plant biomass, but negatively correlated to the stability individual species’ biomass. McNaughton (1997) also found that more diverse plots in African grasslands showed greater variation in species richness in response to perturbations than the less diverse ones. Equally, Sankaran and McNaughton (1999) report experimental evidence supporting the hypothesis that diversity promotes instability at the species abundance level. Frost et al. (1995) found that zooplankton biomass was more resistant to acidification than species-level properties. Schindler (1990) also reports that species or population level indicators were more sensitive than ecosystem-level properties in response to acidification and eutrophication in fresh-water ecosystems. Thus, the empirical evidence seems to reveal that increasing diversity beget stability at the community level but instability at the population level. According to Tilman (1996), these apparently contradictory results can be linked, taking into account the effect of inter-species competition (which is larger in more diverse communities). This effect may resolve also the apparent contradictions between empirical and theoretical results. This ‘compensatory’ mechanism predicts that ecosystems will be more robust to perturbation than their components. Thus, environmental services relying on ecosystem-level properties would be more reliable (more stable) than those depending on particular species or populations.

None of the previously mentioned studies show a clearly discontinuous relationship between stability and diversity. Nevertheless, McNaughton (1993) points out that “it is probable that there is a threshold of change that will overcome the damping effect of biodiversity, with an associated break point of ecosystem function to quite different levels”. Thus, he suggests an ecological threshold depending on the magnitude of the perturbation (species deletion). On the other hand, Tilman and Downing (1994) found a convex (logistic-like) curve for the relationship between species diversity and community resistance (measured as total biomass). Although the experimental design of this study has been criticised (Givnish, 1994; Huston, 1997; Chapin et al., 1998), this result may indicate that above a certain diversity level there are no additional increments in ecosystem stability when biodiversity is increased. That is, there seems to be a certain ‘saturation point’. If this pattern can be generalised, then there could be species diversity thresholds below which the system becomes increasingly unstable as diversity decreases. Above this ‘stability’ threshold, the system is robust enough to resist species deletion without undermining its capacity to absorb perturbations. However, with the current knowledge, it is not possible to propose this kind of pattern as a universal phenomenon across different systems, parameters and perturbations.
To study empirically the stability of real ecosystems requires identification of stable equilibria or persistent structures. This is undoubtedly a hard task for ecologists because events occurring during community succession or assembly can lead to important differences in community structure (Connell and Sousa, 1983; Drake, 1991; Dodd et al., 1995). Stochastic events, like the sequence of species invasion, can result in vast differences in community organisation after a perturbation. This makes the notion of equilibrium or attractor state very evasive and difficult to determine in real ecosystems (Jhonson et al., 1996). Because of that, for some authors, to look for a relationship between diversity (or complexity) and stability is pointless (Sutherland, 1981). Due to these inconveniences, despite the relatively abundant ecological literature dealing with this issue, it is still very difficult to arrive at some universal proposition. Perhaps the emphasis on these sweeping generalisations is the consequence of a tradition of analysis in ecology that has been inherited from classical physics (Holling, 1973), which does not take into account in an appropriate way the inherent complexities of the ecological systems and their relevant qualitative properties. On the other hand, for some ecologists the abstract nature of the resiliences concept and the lack of a generally accepted resilience measurement limits its applicability for guiding policy (Orians, 1996; Risser, 1996). From an operational perspective, the two most used connotations of resilience face difficulties because of the complications of identifying stable equilibria. Thus, the common statement in ecological economics, that environmental policies have to be designed intending to maintain ecological resilience, is helpless unless resilience is operationally well defined.

3. Biodiversity and ecosystem function

3.1. Theoretical hypotheses

Environmental services may rely upon the functioning of ecosystems, that is, on aggregated ecosystem processes involved in the flux of mass or energy. The relationship between biodiversity and ecosystem processes is a topic receiving increasing interest in the ecological literature. Vitousek and Hooper (1993) recognise three possible relationships between biological diversity and ecosystem-level biogeochemical functions, (1) linear; (2) asymptotic and (3) non-existent. Case 1 is also called the ‘diversity–stability hypothesis’ (Jhonson et al., 1996; Boucher, 1997). Lawton (1994) makes a classification where case 2 is called ‘redundant species hypothesis’. He also identifies an ‘idiosyncratic’ hypothesis, which proposes a basically unpredictable relationship between diversity and ecosystem function. Another kind of relationship, comprising a range of possible responses, may appear if all species make a contribution in an additive but unpredictable way. This has been called the ‘rivet hypothesis’.

The ‘redundancy hypothesis’ (Walker, 1992; Lawton and Brown, 1993; Walker, 1995) suggests that in terms of ecosystem function, there is a certain degree of substitutability between individual species, especially if they belong to the same ecological functional group. The term functional group refers to species of the same ecosystem, which share common features determining ecosystem functioning and organisation (Schulze and Mooney, 1993). The ‘rivet hypothesis’ (Ehrlich and Ehrlich, 1981; Lawton, 1993) suggests that the function of species in ecosystems can be analogous to the function of rivets in an airplane. Both systems can afford continual extraction of its constituent components without experimenting a loss of function. However, after a certain point this capacity is lost and only one additional species extinction (rivet popped) may cause a collapse in the functional properties of the system. This vision makes emphasis on the unpredictability of this possible threshold. Ehrlich and Walker (1998) note that there is no essential difference between the rivet and the redundancy hypothesis. The rivet hypothesis recognises the existence of redundancy in ecosystems but emphasises the ignorance of which species we can afford to lose. The redundancy hypothesis points out that deletion of some species may have no immediate significant impact on ecosystem function, but redundancy is likely to be key in the long run because it allows the system to have a higher capacity to afford perturbations.
(Grime, 1998). That is, species redundancy contributes to ecosystem resilience (Walker, 1995).

According to the rivet-redundancy hypothesis, a ‘non-linear’ relationship between biotic diversity and ecosystem function may occur (Solbrig, 1993; Naeem, 1998). Starting from this proposition, Carpenter (1996) suggests a pattern for the relationship between diversity and ecosystem processes rate very similar to that schematised in Fig. 1, when ‘biodiversity’ is in the abscissa axis and ‘ecosystem function’ in the ordinate axis (case a). This is clearly a hypothetical case of ecological thresholds with a high degree of uncertainty because the threshold of diversity is basically uncertain with our current knowledge of ecosystems behaviour (Gitay et al., 1996; Levin, 1993).

3.2. Empirical evidence

Although the discussion about the relationship between species diversity and the rate of ecosystem processes started with Darwin, empirical tests intentionally addressing the issue are relatively recent. The debate started to receive considerable scientific attention (Baskin, 1994; Lacroix and Abbadi, 1998) after a laboratory study made by Naeem et al. (1994). They manipulated plant and animal community diversity to create low, intermediate and high diversity microcosms. The rates of five ecosystem processes were measured, community respiration; decomposition; nutrient retention; plant productivity and water retention. They found that CO₂ community consumption and plant productivity were positively related to species richness. For the rest of the processes there were no consistent patterns of variation. Tilman et al. (1996) made a field experiment where plant species diversity was also intentionally manipulated to address its effects on ecosystem function. They found that both plant productivity and resource utilisation were significantly greater at higher diversity levels. The pattern that better described the relationship between productivity and diversity was again a convex (logistic-like) curve. This pattern is consistent with the redundancy hypothesis and it is predicted by mathematical models assuming increasing competition and increasing chance of having better competitors as species richness increases (Tilman et al., 1997b). As it was stated above, a logistic-like curve rela-
A relationship may encompass a certain ‘saturation’ (redundancy?) point (Tilman, 1997). This may imply also a diversity ‘threshold’, below which the system starts to experience increasing variability in its processes (McGrady-Steed et al., 1997) and increasing susceptibility to perturbation (in this case, to species deletion).

### 3.3. Species or functional diversity

Tilman et al. (1997a) made a field experiment, methodologically very similar to Tilman et al. (1996), but this time controlling, in addition to species number, the functional diversity and the functional composition in a grassland plant community. They measured the rate of six ecosystem processes and concluded that functional group diversity may be a stronger determinant of ecosystem processes than species richness, per se. Hooper and Vitousek (1997, 1998) manipulated a grassland plant community to have a gradient of functional groups diversity, controlling also for composition (the identity of the functional groups). They also concluded that functional composition can have a larger effect on ecosystem processes than functional richness, per se. Wardle et al. (1997a) arrived at similar conclusions studying several ecosystem processes along a gradient of island-area (diversity). Wardle et al. (1997b), Nilson et al. (1999) studied the effect of litter composition manipulation and diversity on below-ground processes. They found that the nature of effects tended to be idiosyncratic — dependent on the particular species composition of the litter mixture.

For some authors (Grime, 1997; Wardle, 1999), these studies disclose a debate about the relative importance of biodiversity and species composition (functional types or species identity) in determining ecosystem function. According to Lawton et al. (1998a,b), this debate misses the point because no two plant species within a functional group have identical niches and because functional guilds are arbitrary divisions of continuous niches. Therefore, the results showing that functional composition explains more than species diversity about ecosystems functioning may be tautological, because by definition different functional guilds embrace species groups with large niche differences (Chapin et al., 1997). This implies that functional groups deletion necessarily will produce a great impact on ecosystem processes.

Studies dealing with the relationship between diversity and ecosystem function have been criticised because, according to some authors (Huston, 1997; Hodgson et al., 1998), they do not control in an appropriate way the functional traits of involved species. This does not establish an unequivocal relation between species richness and ecosystem function. However, these criticisms fail to recognise that it is impossible to manipulate species number without changing the species composition of the sample. For environmental policy, the relationship between species number and ecosystem function in abstracto is not as relevant as knowing which species are prone to become extinct, and to have the possibility to forecast the ecosystem-level consequences of their loss. Nonetheless, the ecological science is far from having this predictive capacity and much more research is needed in this area. Some issues where special attention should be paid are, (a) ecosystem consequences of natural extinction; (b) down and up food web effects of changing diversity at one trophic level; (c) the functional effects of diversity diminution on different marine ecosystems (Covich et al., 1999); (d) the ecosystem repercussion of below-ground biodiversity change, including microbial deletion, in terrestrial (Brussaard et al., 1997; Wall, 1999; Wall and Moore, 1999), fresh-water (Palmer et al., 1997) and marine ecosystems (Snelgrove et al., 1997); (e) the ecosystem role of individual species or processes, specially those being prone to be lost by human impacts (Risser, 1995); (f) the effect of different temporal and spatial scales, and the possibility of discontinuities’ occurrence on the relationship between diversity and ecosystem function.

### 4. Examples of multiple states and thresholds

The existence of multiple stable states or alternative equilibria in ecological systems has been recognised since the 1960s in the ecological litera-
ture. Multiple stable points may occur at different levels of the ecological organisation, but usually they are defined as alternative species assemblage in a community (Scheffer et al., 1993) or multiple possible stable densities of a population (Lernardsson, 1994). Alternative assemblage of species may arise due to differences in the historical development of the community, like a different sequence of recruitment (Sutherland, 1974; Drake, 1990), or due to the effects of physical or biological perturbations, like changes in nutrient concentration or species deletion or invasion (Barkai and McQuaid, 1988). The flip from one possible population density to another may also be triggered by perturbations (May, 1977). Very often, the shift between alternative states in ecological systems occurs suddenly and comprises the existence of threshold effects (Wissel, 1984). This section will continue making reference to some of the already described examples of ecological discontinuities as a consequence of anthropogenic perturbations. The current five most common ways of human impacts on natural ecosystems are considered separately.

4.1. Population harvesting

Species abundance diminution as a consequence of harvesting is one of the most common perturbations on ecosystems. Humans can directly take away individuals for consumption (like in a fishery) or may induce indirectly higher rates of mortality in a certain population due to changes in its biological or physical environment. Multiple stable population densities are predicted by one of the simplest predator–prey models. Fig. 2 illustrates one of the possible outputs of this model where alternative states are present. \( P \) represents prey abundance. Following May (1977), the prey population rate of growth \( G(P) \) in the absence of predation follows the solid curve. The rate at which \( P \) changes is given by \( \frac{dP}{dt} = G(P) - C(P) \). The prey’s abundance will tend to reach an equilibrium level when the natural growth rate equals the loss rate. Thus, at intermediate levels of predator’s density, the points A and C correspond to locally stable states, whose domain of attraction are divided by the unstable equilibrium point corresponding to B. For high and low predator’s density, the system will have only one equilibrium (D and E, respectively). If the predator’s abundance is intermediate and the system is in the point C, prey population harvesting can move \( P \) to levels lower than ‘\( T \)’ (threshold). After that, the system may change to another domain of attraction given by A. On the other hand, if the predator’s population is harvested, the rate of consumption may change, for instance, from \( C(P)_2 \) to \( C(P)_3 \) and the system may flip from C to D. This kind of model has been used to conceptualize alternative equilibria in, among others, grazing systems (Noy-Meir, 1975), zooplankton communities (Steele and Henderson, 1981), marine fish populations (Steele and Henderson, 1984), coral reefs species (Knowlton, 1992) and the ecological effects of introduced animals in islands (Sinclair et al., 1998).

Alternative species composition of Alaskan subtidal and intertidal kelp and sea grass beds associated with the presence or absence of sea otters have been one of the most cited examples of community alternative stable states. Otters have been hunted for trade and local consumption. Due to over-harvesting, in some locations they have completely disappeared. Presence of sea otters is correlated to communities with high macrophyte productivity, high fish diversity and density, low invertebrate productivity and high abundance of harbor seals. On the contrary, when sea otters are absent these communities are characterised by low macrophyte productivity, low
fish diversity and density, high invertebrate productivity and low abundance of seals (Estes and Palmisano, 1974). Intermediate states between these two community configurations are reported to be both rare and transitory. Moreover, abrupt transitions between forested and deforested kelp forests have been reported worldwide (Estes and Duggins, 1995). The effects of otter disappearance on kelp forests are equivalent to overfishing in coral reefs, but in the latter case urchins blooms destroy the reef framework (Kaufman and Dayton, 1997).

Although the alternative states in the kelp forests comprise different species assemblage in a community, the above-mentioned population model has been used to explain this phenomenon. Estes and Duggins (1995) propose this kind of predator-prey model to describe the top–down control that otters exercise in kelp forests. In this community, otters are top predators who consume marine invertebrate, mainly sea urchins. Sea urchins graze on macrophyte, which are the most important source of food and shelter for fishes. Finally, fishes are the main food source of harbor seals. Thus, a radical change in the community species assemblage can be triggered by changes in the domain of attraction in predator–prey (otters-urchins) dynamics. In this case, it is a consequence of predator population harvesting. Graphically (in Fig. 2), this would mean, for instance, a change in the prey loss rate from $C(P)_1$ to $C(P)_3$. This involves a drastic increase in the prey (urchins) population (from ‘a’ to ‘d’). This change of attractor encompasses a ‘cascade’ effect.

Interpreting paleoecological evidence in the Aleutian islands, Simenstad et al. (1978) suggest that overexploitation of otters is not a recent phenomena, but even occurred many thousands of years ago by the action of aboriginal Aleuts. This led the nearshore community to alternate between one state dominated by macroalgae and any other dominated by epibenthic herbivores. In response to this switch, human populations had to shift their diet from one based on fishes and marine mammals to an alternative one based on marine invertebrates. This is a clear example of how ecological discontinuities may comprise changes in environmental services. In this case, environmental services disruption was afforded by human population with an adaptive strategy, to shift the diet. Human adaptive strategies may reduce social costs associated with environmental (discontinuous) change. This can be a further complication to the already hard task of estimating social costs of environmental changes, because human adaptive responses to ecological discontinuities are difficult to predict. The adaptive strategy of human societies to cope with environmental change is one of the main arguments for adopting a technological-optimism perspective and to assume substitutability between natural and human-made capital in environmental economics.

The effect of otters on kelp forests is a typical example of keystone species. This term was proposed by Paine (1969) to emphasise the importance of top predators in determining the structure of animal communities by keeping low the density of prey species, which in the absence of the predation would dominate the system. Later, the term was extended to include those low abundance species whose deletion produce important effects on certain traits at the community level (Power and Mills, 1995). Keystone species may exercise great influence on the structure of communities through consumption, physical or biological habitat modification, mutualism, diseases, seed dispersal, resource provision, etc. (Power et al., 1996). Species which exert a great influence on the ecosystem level modulating the availability of resources to other species by physical state changes are also called ecosystem engineers (Jones et al., 1994; Lawton and Jones, 1995). The term keystone species has been useful in emphasising the inequalities in the relative importance of species in maintaining ecosystem function (Grimm, 1995). However, its application in environmental policy and conservation strategies can be less than straightforward because the label ‘keystone’ is relative to a temporal and spatial scale and also to certain assemblage of species and physical conditions, that is, it is context-dependent (Bond, 1993; Navarrete and Menge, 1996). This makes it difficult to confidently define a priori which species, and under
which circumstances, are keystone (Mills et al., 1993; Paine, 1995). However, despite its operational difficulties, the keystone species concept may be very helpful to understand the mechanisms underlying ecological thresholds and discontinuities. It exemplifies very well how relatively small changes in the components of an ecosystem, like harvesting on a keystone species, may involve sudden and far-reaching shifts in the system’s properties.

Overharvesting of natural populations may also influence natural populations cycles, making less predictable and more frequent natural fluctuations (Steele and Henderson, 1984). These fluctuations involving dramatic, uncertain and rapid changes in population abundance, although not equivalent to shifts between alternative states because they are transitory, may encompass important economic costs. Classic examples of abrupt population abundance changes linked to overharvesting of fish populations are the sardine stock collapse in California and Japan in the late 1940s and the anchovy disappearance in Peru and Chile in the 1970s (Botsford et al., 1997).

4.2. Pollution

Changes in the physical environment of ecosystems may produce a switch between alternative equilibria. These cases are proposed as examples of alternative stable states because once the switch has already occurred, it is not enough to recover the initial physical conditions to return the system to the original state. One of the most cited examples of ecological discontinuities linked to changes in nutrients concentration is the existence of alternative equilibria in temperate shallow lakes. It has been reported that shallow lakes in temperate zones can present two basic organisations, a clear state dominated by aquatic vegetation (macro-algae) and a turbid state characterised by high algal (phytoplankton and cyanobacteria) biomass (Weisner et al., 1997). Continual nutrient loading to clear lakes can be afforded by the system until a certain nutrient concentration level is reached (threshold). After this point, a shift to the turbid state is highly probable. Once the system has switched to a turbid state and cyanobacteria become the most important primary biomass, it takes a strong nutrient reduction (much below the value of the initial threshold) to enable macrophyte recolonisation (Moss et al., 1996; Scheffer et al., 1997). This phenomena, as it is suggested by Scheffer et al. (1993), can be conceptualised graphically when the variable ‘nutrient concentration’ is in the abscissa axis and ‘algal biomass’ in the ordinate axis in Fig. 1 (case b). Thus, this is clearly a case where ecological thresholds and discontinuities are present. Allelopathic substances and habitat conditions encouraging predator development or avoiding light penetration are some of the proposed synergetic mechanisms to explain the stability of these alternative states. Due to the complexity of the biological interactions underlying this pattern, and because the switch from the clear to the turbid state can be mediated by stochastic events like hurricane winds (Bachmann et al., 1999), the precise value of the nutrient concentration threshold can be hardly predicted (Donabaum et al., 1999).

Thus, in this relatively well-studied case of alternative equilibria, the thresholds values are rather uncertain. In this example, ecological discontinuities are associated with important loss of environmental services, mainly fishery and recreation.

4.3. Habitat fragmentation

Habitat fragmentation, the process of subdividing a continuous habitat into smaller pieces, is the major cause of species extinction (Tilman et al., 1994a,b). Currently, it is probably the most important way of human intervention on natural ecosystems. The existence of thresholds associated to the habitat size is well recognised in the ecological literature. The theory of island biogeography (MacArthur and Wilson, 1967), based on certain assumed behaviour of immigration and extinction rates, predicts a linear and continuous relationship between island size and species richness. However, Ward and Thornton (1998) suggest that there may be alternative equilibria at intermediate island sizes. This means that islands of the same (medium) size may have alternative equilibrium species numbers. This model predicts a relationship between island species richness and island
size similar to Fig. 1, when the variable ‘equilibrium species number’ is in the ordinate axis and ‘island size’ in the abscissa axis (case c). This may comprise a sudden ‘switch’ between species-rich to species-poor habitats when island (patch) size is reduced below a certain threshold. Patches in a fragmented habitat can be considered ecologically equivalent to islands in a sea. Thus, these predictions also could be applied to the general relationship between species richness and habitat size. Actually, Metzger and Décamps (1997) propose a similar ‘non-linear’ and discontinuous relationship between habitat fragmentation and biodiversity. This relationship can be summarised in Fig. 3, when the variable ‘proportion of destroyed habitat’ is in the abscissa axis and ‘biodiversity’ in the ordinate axis (case a). This model is based on two basic theoretical developments about the ecological consequences of habitat fragmentation.

1. One is related to habitat’s geometrical qualitative and quantitative changes along a gradient of fragmentation. Spatially explicit models predict the existence of thresholds determining discontinuities in the relationship between the proportion of the original habitat destroyed and landscape connectivity, the spatial contagion of habitats (With and Crist, 1995). These discontinuities arise because quantitative changes in the original habitat size lead to qualitative shifts in the properties of the patches. Bascompte and Solé (1996) represent these non-linear interactions in a graph similar to Fig. 3, where the variable ‘largest patch size’ is in the ordinate axis and ‘proportion of destroyed habitat’ is in the abscissa axis (case b). Hence, after a certain fragmentation threshold, the effect of destroying an additional portion of the habitat is no longer quantitative (the reduction of the largest-patch’s size), but qualitative, the original habitat starts to be broken into smaller patches. After this threshold, the ‘border effect’ and isolation might strengthen the effect of habitat loss and the decline in population size will be faster than predicted by a simple diminution in the ‘sample’ area (Andrén, 1994). Population responses to changes in habitat size can vary greatly from one species to another (Diamond, 1982; Soulé et al., 1992; Kruess and Tscharntke, 1994; Tilman et al., 1994a,b; Lawton et al., 1998a,b). Because they are always referred to a species and to a certain spatial and temporal scale, to make these thresholds operational (predictable and measurable) for environmental policy embraces serious difficulties.

Fig. 3. Habitat fragmentation and ecological thresholds.
2. The other theoretical development predicting ecological thresholds linked to habitat fragmentation are metapopulation models, which propose multiple equilibria in the relationship between the probability of species occurrence and the area and isolation of the patches (Hanski and Gyllenberg, 1993). Graphically, this relationship, which has been supported empirically, can be visualised in Fig. 1 (case d), when the variable ‘fraction of occupied habitat’ (in a patch network) is in the ordinate axis and ‘potential colonisation rate’ (which depends on the patch size and the distance between patches in the network) in the abscissa axis (Hanski et al., 1995). At intermediate values of potential colonisation rate, small changes in patch size may cause the extinction of the metapopulation. This suggests the existence of an extinction threshold, which can be defined as the minimum proportion of suitable habitat that is necessary for population persistence (Lande, 1987). Extinction thresholds can be visualised in Fig. 3 (case c), when the variable ‘proportion of suitable territories occupied’ by a certain species is in the ordinate axis and the ‘proportion of destroyed habitat’ in the abscissa axis (Andrén, 1996). The results of some empirical studies are compatible with these propositions (Didham et al., 1998).

Extinction thresholds are produced by changes in the interaction between individuals and the surrounding landscape, but also by changes at the population or genetic level. At these levels there also can be discontinuities. Kuussaari et al. (1998) suggest that from certain population size down, the Alle effect (a diminution in population growth rate at low population densities) can play an important role in promoting extinction. Frankham (1995) reports empirical evidence showing that the most likely relationship between inbreeding (the probability to mate with relatives) and extinction (due to genetic reasons) is a threshold relationship. On the other hand, at higher organisational levels, system feedback and synergism may be involved in the appearance of discontinuities. As it was stated previously, species richness decline or keystone species loss produced by habitat fragmentation may lead to further species extinction due to ecosystem function or stability loss. This may generate a cascade effect and a switch in ecosystem properties.

4.4. Ecosystem management

Human intervention on natural ecosystems through manipulation of key internal or external variables can be also a source of ecological discontinuities. Human management of ecosystems usually is referred to as control of ecological perturbations. Manipulation of natural perturbations in grasslands is one of the cases where the role of ecosystem management in generating discontinuities has been addressed. Temperate savannas around the world have experienced very often rapid switches from grass-dominated to woody-plants-dominated communities (Fuhlendorf et al., 1996). Overgrazing, fire suppression and CO₂ concentration rise have been cited as the most likely proximate causes of woody plant invasion into grasslands (Archer, 1989). All of these causes are related to human activities.

The most used rangeland management approach assumes that there is a ‘climax’ state in grasslands. According to this approach, any transformation in the properties of the system due to human perturbation may be reversed simply by suspending the perturbation (Laycock, 1991). Nevertheless, observations reveal that once the system has switched from grassland to woodland due to human management, stopping the perturbation during many years usually is not enough to recover the grass-dominated state (Walker et al., 1981). Once woody plants have substituted grass cover, positive feedback effects make the woodland system highly robust (Perrings and Walker, 1997). This led some authors to propose an alternative grassland dynamics model, which assumes the existence of alternative stable states and a non-linear relationship between perturbation intensity or frequency and grassland condition. This model makes explicit the existence of ecological discontinuities and thresholds (Walker et al., 1981; Archer, 1989; Fuhlendorf et al., 1996). Above a certain critical value of grazing pressure or below a certain fire frequency the system flips from the grassland to the woodland state. In
order to recover the initial state, it is not enough to reduce grazing intensity or to increase fire frequency to the original threshold level. Reversion may require strong intervention by the manager, such as the application of herbicides or heavy machinery (Friedel, 1992). This dynamic can be visualised in a graph similar to Fig. 1 (cases e and f).

The case of fire reveals how management strategies intending to diminish variability may actually generate greater instability in the system. Suppression of forest fire in the national parks of the US has successfully reduced fire probability. Nonetheless, simultaneously this management has also produced fuel accumulation in the form of biomass, making the system prone to suffer fires of a magnitude never experienced before (Holling, 1986). After this kind of perturbation, the forest may switch to a different state, becoming a grassland for example. Another example, where human management intending to reduce variability in ecological systems may have produced the opposite results, is the control of the spruce budworm in boreal forests. This system experiences a natural cyclic behaviour. Periods of forest biomass accumulation and budworm low abundance are interrupted by periodic (20–40 years interval) and sudden outbreaks of the budworm density, which produces drastic diminution in foliage cover (Ludwig et al., 1978; Holling, 1992). This is a case of ecological thresholds in natural conditions. In order to preserve stable forest conditions, the pulp and paper industry has used insecticides to maintain low budworm density. This strategy has made the system prone to suffer disruptions due to pest outbreak in a magnitude never experienced before. As in the case of fire in forests, this extent of perturbation may switch the system to an alternative state difficult to reverse (for instance, from a conifer-dominated to an aspen-dominated forests).

Another case of natural cyclic pest outbreak occurs in larch ecosystems in central Europe. In this system, the larch bud moth rapidly increase to epidemic levels at intervals of 9–10 years and cause extensive diminution in larch foliage cover (Woodward, 1993). According to Holling (1986), management strategies that do not take into account the role of natural fluctuations in the maintenance of ecosystem properties may generate ecosystem resilience loss in the longterm. However, it is not clear the extent to which this kind of cyclic population behaviour is present in natural ecological systems. According to Holling (1992) cyclic population dynamics are common in nature. However, for Pimm (1984) ‘such dynamics are theoretically fascinating but seem to characterise only a few populations’.

4.5. Biological invasions

Human activity has accelerated the rate of species invasion to new habitats by many orders of magnitude. Only in Canada, around 20% of current plant species were introduced by humans from elsewhere (Vitousek et al., 1997). Biological invasions are probably the second most important way of human-driven species extinction, after land use change (D’Antonio and Vitousek, 1993). Vitousek (1990) points out that biological invasions can alter ecosystem function by modifying three basic properties at the community level (a) resource or water flows; (b) the trophic structure and (c) the disturbance regime. The ecosystem effects of species invasion may be ecologically equivalent to the role of keystone species, exemplifying that the properties of individual species matter at the level of the whole ecosystem (Vitousek and Walker, 1989). Equally, they may comprise feedback effects and sudden shifts in ecosystem properties, although the notion of ecological thresholds has been less studied in the case of biological invasions. An example of the effects on nutrient cycling of invading species is the nitrogen-fixing tree *Myrica faya* in Hawaii. This plant, originally brought from the Azores and Canary Islands to Hawaii in the 19th century, possesses a nitrogen fixing actinorrizal symbiont which is absent in the native vegetation in Hawaii. Vitousek et al. (1987) determined that the presence of Myrica quadrupled the amount of nitrogen entering certain sites in Hawaii and increases the overall biological availability of nitrogen. This clearly alters ecosystem level characteristics, although the longerterm consequences of this alteration remains to be determined. In general,
nitrogen addition in plant communities has produced a diminution of species richness (Vitousek, 1994; Tilman, 1997). Species richness changes, as it was stated above, can produce a cascade effect and discontinuities at the ecosystem-properties level.

The construction of the Erie and Welland canals in the 19th century induced an unexpected invasion of sea lamprey in the Great Lakes, which after 40 years produced a drastic abundance diminution in most of the economically important fish species. Lamprey invasion comprised substantial economic costs due to fishery and recreational activity decline. Predation was the main mechanism by which lampreys affected the structure of the Great Lakes fish community (Aron and Smith, 1971). In this sense, this example is equivalent to the previously mentioned sea otter case in Alaska. This constitutes an example of how invasive species may cause major and sudden disruptions in environmental services by altering the community trophic structure.

The effects of grass invasions on the fire regime are a worldwide well-documented event. The increase of fire frequency as the consequence of alien grass invasion has been reported in Latin America, North America, Australia and Hawaii (D’Antonio and Vitousek, 1992). Many grass invasive species re-grow more quickly than native plants after fire. Moreover, they are more flammable and have fire-resistant seeds. Due to these properties, grass invasions may initiate a grass/fire cycle, where invading grasses promote fire, which in turn favours foreign grasses over native species (Hughes et al., 1991). This cycle may have important consequences at the ecosystem level and it can produce higher incidence of fire and a relatively rapid switch from woodland to grassland state. Actually, it has occurred in many parts of the world (D’Antonio and Vitousek, 1992). Undoubtedly, this could imply dramatic shifts in local ecological services and considerable economic costs.

5. Concluding remarks

There are multiple theoretical developments in ecology predicting the existence of ecological discontinuities triggered by threshold values of internal or external variables to the ecological systems. Empirical studies reveal that ecological discontinuities as the consequence of human impacts are not uncommon in nature. Nevertheless, these studies face some difficulties, mainly because the definition of alternative stable states is highly dependent on the chosen temporal and spatial scales, as well as on the adopted notion of attractor. These difficulties also impose problems to the two most common notions of resilience, a concept that has not received a very rigorous treatment in the ecological economics literature. For many systems, the causes triggering ecological discontinuities are relatively well known. However, at the current state of knowledge, the ecological science is more able to predict the magnitude of change (the possible alternative state) than the threshold values. As prices are equally not able to assess the proximity of a system to a discontinuity, it seems that neither specialists nor consumers are capable of predicting exact ecological thresholds. Further ecological research probably will improve this predictive capacity, but given the complexity of the phenomenon and the stochastic variables involved, large uncertainties are likely to remain. To develop methods to deal with these uncertainties is a compelling task for policy-makers and a big challenge for participatory decision-making processes.

References


