Natural Science Constraints in Environmental and Resource Economics

Method and Problem

Stefan Baumgärtner

Alfred-Weber-Institute of Economics,
University of Heidelberg, Germany

University of Heidelberg Publications Online
http://archiv.ub.uni-heidelberg.de/Englisch/index.html

Stefan Baumgärtner
University of Heidelberg
Alfred-Weber-Institute of Economics
Bergheimer Str. 20
D-69115 Heidelberg, Germany
baumgaertner@uni-hd.de

This text is the revised version of a Habilitation Thesis that was submitted to the Faculty of Economics and Social Studies, University of Heidelberg, Germany on 06 July 2005 and was accepted on 26 April 2006 (reviewers: Prof. Malte Faber, Prof. Hans Gersbach).

Contents

Acknowledgements

Interdisciplinary research means to bridge the gaps between traditional academic disciplines. This is hardly possible for a single researcher. As this study aims to integrate perspectives, methods and insights from different natural and social sciences as well as the humanities, I naturally owe a lot to many scholars from different academic disciplines – such as economics, ecology, engineering, physics, mathematics and philosophy – who have contributed in one way or another to this research.

Chapters 5, 6 and 10 of this study originated from joint research with Christian Becker, Karin Frank, Frank Jöst, Birgit Müller, Martin Quaas, Jakob de Swaan Arons and Ralph Winkler. Working together with every single one of them was a great opportunity, an even greater pleasure, and indispensable for the kind of interdisciplinary research which essentially makes this study.

A number of people have spared some of their time and discussed drafts or presentations of one or more of the papers, upon which this study draws. I am grateful for their critical and constructive comments, as well as stimulating discussions:

- Geir Asheim, Scott Barrett, Carl Beierkuhnlein, Lucas Bretschger, Erwin Bulte, Jon Conrad, John Coulter, Salvatore di Falco, John Ehrenfeld, Malte Faber, Karin Frank, John Harte, Sönke Hoffmann, Frank Jöst, Giselher Kaule, Bernd Klauer, Andreas Lange, Karl Eduard Linsenmair, Michel Loreau, Reiner Manstetten, Ulf Moslener, Klaus Nehring, Richard Norgaard, Charles Perrings, Thomas Petersen, Rüdiger Pethig, Steve Polasky, John Proops, Clemens Puppe, Inge Røpke, Martin Quaas, Till Requate, Joan Roughgarden, Johannes Schiller, Felix Schläpfer, Armin Schmutzler, Thomas Schulz, Irmi Seidl, Thomas Sterner, John Tschirhart, Frank Wätzold, Hans-Peter Weikard, Gerhard Wiegleb, Bruce Wilcox, James Wilen, Ralph Winkler, Christian Wissel, Anastasios Xepapadeas, David Zilberman;

- Conference, workshop and seminar participants at Aachen, Ascona, Bad Honnef, Berkeley, Berlin, Budapest, Cambridge, Castelvecchio, Chemnitz, Cottbus, Darmstadt, Delft, Geneva, Göttingen, Großbothen, Halle, Heidelberg, Karlsruhe, Leipzig, Lisbon, Mannheim, Monterey, Montreal, Oberflockenbach, Paris, Rethymno, Santiago de Compostela, Schloss Wendgräben, Sousse, St. Andreasberg, Stockholm, Strasbourg, Tenerife, Tutzing, Venice, Versailles, Washington DC, Vienna and Zürich;

- Anonymous reviewers of the journals *Ecological Economics*, *Economic Theory*, *Ecosystem Health*, *Environmental and Resource Economics*, *Journal of Economic Dynamics and Control*, *Journal of Environmental Economics and Management*, *Journal of Industrial Ecology*, *Natural Resource Modeling* and *Resource and Energy Economics*;

The University of Heidelberg's Interdisciplinary Institute for Environmental Economics provided an ideal platform for conducting this research. I am grateful to the directors of the institute – Malte Faber and Till Requate in particular – as well as to all researchers, students and visitors for creating an environment of intellectual openness and scientific excellence. Eva Kiesele and Frank Schwegler have provided valuable research assistance; Dale Adams and Maximilian Mihm have helped improve language and style; Simone Bauer has helped editing the references.

The initial drafts of Chapters 8, 9 and 11 were written while I was a Visiting Scholar with the Energy and Resources Group at the University of California, Berkeley, in the academic year 2001/2002. I am grateful for their hospitality – Dick Norgaard's in particular – and for the most stimulating research environment at Berkeley.

Last, not least, I am grateful for financial support from the Deutsche Forschungsgemeinschaft (DFG) under grant BA 2110/1-1 and from the Volkswagen Foundation under grant II/79 628.

1. Introduction

Most, if not all, environmental problems of our time have their origin in human economic activity. For example, the production of electricity from fossil fuels depletes the Earth's fossil fuel deposits and pollutes the atmosphere with greenhouse gas emissions; the use of water for industrial or agricultural production and as a medium to dispose of wastes causes pollution of surface waters and groundwater reservoirs; the consumption of a huge variety of products leaves behind enormous amounts of wastes, some of which are harmful to human health and ecosystems; the satisfaction of mobility needs by modern traffic systems and infrastructure destroys natural landscapes and habitat for many biological species. In all these cases, the economic benefits, which are the primary justification for action, are intimately linked to environmental problems.[1]

In order to better understand how environmental problems arise from economic activity and how they may be solved, one needs to combine scientific expertise from the natural sciences and from economics. For, it is the domain of the natural sciences to analyze 'nature', while economics studies 'the economy'. In this study, I contribute to this interdisciplinary task in a threefold manner:

1. In Part I, I employ concepts and methods from thermodynamics in order to study how this natural science puts constraints on the transformation of energy and matter in the economic process of production.

2. In Part II, I analyze the problem of biodiversity loss and conservation by combining concepts and methods from ecology and economics.

3. An underlying interest throughout this study is the methodological question of how to integrate concepts and methods from the natural sciences, such as thermodynamics or ecology, and the social sciences, such as economics. This question is discussed in detail in this introductory chapter, in order to provide a methodological basis for the actual analysis in Parts I and II.

[1] In conceptual terms, the structural cause behind many modern-day environmental problems is *joint production* (Baumgärtner et al. 2006). This captures the phenomenon whereby human action always entails unintended side-consequences.

This chapter is organized as follows. Section 1.1 opens the methodological discussion by characterizing the economic approach to studying economy-environment interactions. Section 1.2 discusses various concepts of *nature* in economics, in order to shed more light on this defining element of environmental and resource economics. Section 1.3 then clarifies the role of the *natural sciences* for environmental and resource economics (Section 1.3.1) and addresses the challenge of *interdisciplinary integration* of economics and the natural sciences (Section 1.3.2). It introduces a fundamental distinction between two approaches to incorporating natural science constraints into environmental and resource economics – *method-orientation* and *problem-orientation* (Section 1.3.3). Furthermore, it justifies the focus of this study on *conceptual analysis* (Section 1.3.4).

After the methodological basis is thus prepared, the contents of this study is introduced in the remainder of the chapter. Section 1.4 previews Part I, which deals with thermodynamic analysis of economy-environment interactions and is characterized by method-orientation. Section 1.5 previews Part II, which deals with biodiversity loss and conservation and is characterized by problem-orientation.

1.1 Economics and the Study of Economy-Environment Interactions

According to a classic definition, *economics* is 'the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses' (Robbins 1932: 15). This definition has a wide scope and, consequently, economics approaches a wide range of issues. One of these issues is the relationship between human economic activity and the natural environment, which is the subject of the sub-discipline of environmental and resource economics (e.g. Baumol and Oates 1988, Dasgupta and Heal 1979, Hanley et al. 1997, Hartewick and Olewiler 1998, Kolstad 2000, Siebert 2004 and Tietenberg 2003). In line with Robbins' definition, the approach of environmental and resource economics to studying economy-environment interactions is characterized by

(i) a distinction between *means* – e.g. labor, capital, natural resources, ecosystem goods and services, or the environment's absorptive capacity for pollutants and wastes – and *ends* – e.g. maximizing a firm's profit or social welfare;

(ii) the idea that means such as natural resources are *scarce*, which is usually taken to mean that obtaining and utilizing them carries (opportunity) costs (e.g. Debreu 1959: 33, Eatwell et al. 1987);[2] and

(iii) the existence of *alternatives* in using means to achieve ends, which implies that there is scope for making choices and, at the same time, choices have to be made about how to best use scarce means. *Choice*, thus, becomes the true substance matter of economics.

This characterization has lead to the understanding that economics, including its sub-field of environmental and resource economics, is essentially about optimization under constraints, with an objective function representing ends and constraints as an expression of scarcity of means.

The aspect of scarcity allows defining more clearly the field of environmental and resource economics as a sub-discipline of general economics: environmental and resource economics studies those areas of optimizing human behavior subject to constraints where constraints are imposed by *nature* (Fisher 2000: 189). Examples include the limited stock, concentration and spatial distribution of mineral resources; the natural growth and interaction of biological resources; the diffusion, transformation and decay of a pollutant in an environmental medium; etc. In this view, the laws of nature captured by the environmental natural sciences, such as physics, hydrology, biology, ecology, geology, etc., are necessary for environmental and resource economics to gain an adequate representation of relevant constraints.

This logic justifies an interdisciplinary cooperation between economics and the natural sciences in the study of economy-environment interactions. But there are at least three fundamental methodological problems for any such cooperation (Becker and Baumgärtner 2005: Section 3):

1. The concepts and methods employed in different disciplines of the natural and social sciences – such as thermodynamics, ecology, or economics – stem from, and are shaped by, very different disciplinary traditions, cultures and self-images. It is not at all obvious that they are compatible with each other.

2. Different disciplines have different research interests. What constitutes an 'interesting' question for one discipline may be completely irrelevant for another discipline, even within a common substantive domain, such as 'economy-environment interactions'.

3. It is not even obvious whether different disciplines could agree on the exact substantive content of 'economy-environment interactions'. This requires an answer to the questions 'What exactly is nature?' and 'What

[2] For a more detailed discussion of the concept of *scarcity* of natural resources, goods and services, see Baumgärtner et al. (in press).

exactly is the economy?' Even worse, these questions may not have a precise and unique answer even within each discipline.

As this study takes an economic approach to analyzing economy-environment interactions, the second problem is obviously solved and the third problem reduces to the question 'What exactly is "nature" in the view of economics?'. In the following section, I will address this question in detail. Section 1.3 then addresses the first problem.

1.2 Concepts of Nature in Economics

As the concept of *nature* is crucial for defining the field of environmental and resource economics as a sub-discipline of general economics, one needs to address the question: 'What exactly is "nature" in the view of economics?' As economics deals primarily with the economy, not with nature, one should not expect economics to have a clearly defined and encompassing notion of 'nature'. Yet, there exist a number of different, mostly implicit, notions of nature in economics (Becker 2005, Biervert and Held 1994, Schefold 2001). They are embedded in different perspectives on the relationship between nature and the economy, some of which are discussed in the following.[3] Each of these perspectives highlights a particular aspect of this relationship and, thus, expresses a particular concept of nature.

1.2.1 Nature as Part of the Economy

Natural resources and services have an obvious economic dimension insofar as they may serve as production factors or directly as consumption goods. Examples include the utilization of coal and iron ore for the production of steel, or the appreciation of nature's beauty by tourists during their vacation. In their function as production factors or consumption goods, natural goods share to some extent the general characteristics of any economic good: they are relatively scarce, substitutable against other natural or manufactured goods, subject to subjective valuation, and subject to individual or collective allocation decisions.[4]

In conceptualizing nature as a *set of goods and services* which share the essential characteristics of any other economic good, nature is seen as part

[3]This classification follows Becker (2005) as far as the conceptualization of nature as part of the subject matter of economics is concerned ('what to explain'). In addition, Becker (2005) also discusses concepts of nature as a model for understanding the subject matter of economics ('how to explain').

[4]One notable difference, which sets natural goods apart from other economic goods, is that they are often public goods and that property rights are often not well defined.

of the economy, like any other sector of the economy. It, thus, falls into the domain of economic decision making. This is the traditional understanding of *environmental and resource economics*, as it has emerged in the early twentieth century as a sub-discipline of general economics (Gray 1913, 1914, Hotelling 1931, Pigou 1912, 1920), based on the methodological foundation of neoclassical economic theory. It puts the economic decision maker at center stage, and this procedure also defines 'nature': the economic perception of nature is reduced to those objects and services, and the respective dimension of their physical existence, that are of value to economic agents.

1.2.2 Nature as a Limit to Economic Activity

When people became aware of the existence of global and long-term environmental problems in the second half of the twentieth century – such as depletion of the stocks of mineral resources and fossil fuels, land degradation, overfishing of the oceans, climate change, rupture of the ozone layer, biodiversity loss, etc. – this challenged the view that natural goods and services are essentially like any other economic goods and services, and which stresses the manageability of, and human control over, nature. In contrast, it now became apparent that nature may impose limits to economic activity and, to a considerable extent, is beyond human control and management.[5] The field of *ecological economics* emerged in the 1960s and 1970s from the insight that the traditional approach of environmental and resource economics, which considers particular environmental goods and services that share the general characteristics of any economic good, was too narrow, and the treatment of nature in the analysis of economy-environment interactions needed a systematic and more encompassing approach (Costanza 1989, 1991, Costanza et al. 1997c, Røpke 2004).

A corner stone in the early arguments of ecological economics is the claim that the laws of thermodynamics, which fundamentally govern the transformation of energy and matter, also govern economic action and economy-environment interactions insofar as these consist of energy/matter-transformations. According to the laws of thermodynamics, energy and matter cannot be created or destroyed (First Law), and in any transformation of energy and matter a non-negative amount of entropy is created (Second Law). This should fundamentally constrain the set of feasible economic actions (Ayres 1978, Ayres and Kneese 1969, Boulding 1966, Faber et al. 1995[1983], Georgescu-Roegen 1971, Kneese et al. 1972). More recently, a similar line of argument emphasizes the role of ecological relationships for the functioning and resilience of ecosystems and, thus, their 'carrying capacity' in terms of economic use and impact that

---

[5]This view led to the report on *The Limits to Growth* by the Club of Rome (Meadows et al. 1972). As a matter of history, the view that nature is beyond human control and acts as a limit for human action, governs all pre-modern thinking. It also shows up in the writings of the classical economists, e.g. Malthus (1798).

ecosystems can withstand before they loose their ability to generate ecosystem goods and services (Arrow et al. 1995, Daily 1997b, Gunderson and Holling 2002, Perrings 1995b, 2001, Perrings et al. 1995a).

The conceptualization of nature implicit in these arguments is that of a *set of laws of nature* which systematically determine what is possible and what is not – not only in the realm of the natural phenomena, but also as far as economic action and economy-environment interactions are concerned. Thereby, nature limits the potential scope of economic action. It is not only relatively scarce, as in the view of environmental and resource economics, but imposes an absolute scarcity on the economy (Baumgärtner et al., in press). This conception of nature also leads to a modified view on the human economy, which appears to be limited by, and contingent upon, nature.

1.2.3 The Economy as Part of Nature

One intellectual consequence of recognizing nature as a limit to economic action is a fundamental change in perspective: nature is no longer seen as part of the economy but the economy is seen as part of nature.[6] Accordingly, the 'vision' (in the sense of Schumpeter)[7] of *ecological economics* is that the human economy is an open subsystem of the larger, but finite, closed, and non-growing system of non-human nature (e.g. Ayres 1978, Boulding 1966, Daly 1991[1977], Faber and Proops 1998, Georgescu-Roegen 1971). In this perspective, the dynamics of economy-environment interactions appears as a co-evolution of two systems that both have their internal structure and dynamics, and mutually influence each other's development (Norgaard 1981, 1984, 1985, 1994).

This view of the economy as part of nature is very encompassing. It includes different aspects of the economy-environment-relationship under a unifying perspective, which appear isolated in the nature-as-part-of-the-economy-perspective and the nature-as-a-limit-to-economic-activities-perspective respectively:

- First, nature provides a number of goods and services that may be of value for, and utilized by, optimizing economic agents. This is the aspect of the economy-environment-relationship which has been stressed by the nature-as-part-of-the-economy-perspective (Section 1.2.1) and which is the focus of traditional environmental and resource economics.
- Second, while these natural goods and services share essential characteristics of other economic goods, they are crucially distinct from the latter

---

[6]Brown (2001: 5) argues that the change of perspective from nature-as-part-of-the-economy to the-economy-as-part-of-nature amounts to a scientific revolution not unlike the transition from the geocentric to the heliocentric world view in the Copernicanean revolution.

[7]Schumpeter (1954: 42) defines a *vision* as the 'preanalytic cognitive act that necessarily precedes any scientific analysis'.

in that they come from a geobiophysical environment governed by laws of nature which are beyond human control and, thus, impose exogenous (and, in general, not constant) limits to human economic activity. This is the aspect of the economy-environment-relationship which has been stressed by the nature-as-a-limit-to-economic-activities-perspective (Section 1.2.2) and which was the focus of early ecological economics.

Thus, nature is conceptualized as a *set of goods and services plus a set of laws of nature* governing the provision of these natural goods and services.

1.3 Methodological Position of this Study

1.3.1 The Role of the Natural Sciences

In this study, I adopt the perspective of the economy as part of nature (Section 1.2.3), since this is an encompassing economic perspective on the relationship between human economic activity and nature. As described above, in this perspective nature provides a number of goods and services which share essential characteristics of other economic goods, but are crucially distinct from the latter in that they come from a geobiophysical environment governed by laws of nature.[8]

Within this perspective, the role of the natural sciences now becomes clearer. The natural sciences, such as physics, chemistry, biology, hydrology, geology etc., are necessary for the field of ecological, environmental and resource economics to the extent that their concepts and laws give a clear, systematic and encompassing description of the characteristics of the goods and services provided by nature, and of the relationships that govern their generation and provision. Thereby, they describe the natural world as offering a potential for, but also as setting limits to, economic action.

The relationship between the natural sciences and economics in the study of economy-environment interactions then is, in principle, as follows. Concepts and laws of the natural sciences are essential ingredients in the characterization and delimitation of the 'commodity space' (as far as natural goods and services are concerned) and the 'set of feasible economic actions'. That is, they serve to conceptualize the *objects* of economic action (as far as natural goods and services are concerned) and to formulate the *constraints* on economic action (as far as they are imposed by laws of nature). Their role is limited to this particular task. The ranking of feasible actions and the explanation of which action is chosen by an optimizing economic agent (*homo economicus*) do not

[8]Being based on such a perspective, this study cannot be classified as belonging either to environmental and resource economics or to ecological economics. Instead, it displays typical characteristics of both approaches.

require any input from the natural sciences, but are subject to economic analysis proper.[9] Thereby, ecological, environmental and resource economics is, first of all, economics in that it centrally studies optimizing human behavior under constraints based on the standard concepts and tools of economics; it is informed by the natural sciences insofar as the formulation of the commodity space and constraints is concerned.

1.3.2 The Challenge of Interdisciplinary Integration

Insofar as ecological, environmental and resource economics is defined as a subfield of general economics by the integration of laws and concepts from the natural sciences, it is inherently interdisciplinary. Hence, the methodological challenge arises of how exactly to integrate concepts and laws from the natural sciences into an economic analysis. Different procedures and degrees of interdisciplinary integration are imaginable and have been distinguished (e.g. Becker and Baumgärtner 2005: Section 3.2):

(i) One potential approach of how to study a subject matter from different disciplinary perspectives is a **multidisciplinary analysis** in which different disciplines make statements about the same subject matter, but they do so in isolation. That is, each discipline addresses the aspects that it considers relevant, and it does so in its own terminology and based on its own set of concepts, methods and theories. For example, in a multidisciplinary analysis of greenhouse gas emissions by economists, legal scientists and atmospheric scientists, the economists would study the optimal allocation of emissions based on their costs and benefits; the legal scientists would study the restrictions on emissions imposed by existing, national or international, regulations; and the atmospheric scientists would study the physical or chemical impact of emissions on the state of the atmosphere. Their results would typically be reported as an additive compilation of independent disciplinary sub-reports, each written by one disciplinary sub-group. In such a multidisciplinary analysis, the different disciplinary contributions are not integrated in any substantial manner; this task is left to the reader. The question remains open whether the different disciplines have really studied the same subject matter.

(ii) A method may be transferred from one discipline into another one, where it is then applied to the substantive domain of the importing discipline according to the scientific criteria and organizational structures of that discipline. Such a **transfer of method** is a unidirectional relation and

[9] In contrast, some have argued that seeing the economy, and human action in general, as fundamentally constrained by nature should make a difference for how to describe and analyze human behavior (e.g. Costanza et al. 1997a, Daly 1991, 1992a, Faber et al. 1996).

does not aim at a bidirectional relationship between two disciplines. For example, economics has adopted the so-called 'Le Chatelier Principle' from classical thermodynamics and uses it (under the name of 'comparative statics') to study the properties of economic equilibria (Samuelson 1947).

(iii) In an **interdisciplinary division of labor** different disciplines address the same subject matter in such a manner that they each base their investigation on their own disciplinary set of concepts, methods and theories, and exchange results via clearly defined data interfaces. This may be a recursive procedure. In this approach the interdisciplinary coordination and cooperation pertains to the input and output of data and results; it does not cover the internal elements and structure of the disciplinary analyses. An example is the interdisciplinary analysis of global anthropogenic climate change by coupled simulation models, where demographic and economic models produce projections about future emission paths; these serve as input into climate models, which predict climate change; and the climate data thus obtained are then, again, fed into the economic models of optimal emission choice.

(iv) While in an interdisciplinary division of labor each discipline retains autonomy over how to set up and carry out its analysis, a closer coordination and cooperation is possible. In an **interdisciplinary integrated analysis** the concepts, methods and theories of different disciplines are closely related and adjusted to each other with regard to the joint interdisciplinary scientific aims. This happens in a discussion process among scientists that clarifies what disciplinary concepts, methods and theories are adequate to the joint interdisciplinary endeavor, how they relate to each other, and how they need to be adjusted to each other with regard to the interdisciplinary scientific aims. An example is the ecological-economic model analysis of biodiversity management policies, where the different disciplinary sub-models are adjusted to each other within a joint research perspective, so that they are formulated in the same functional terms and operate on similar spatial and temporal scales (Wätzold et al., forthcoming).

(v) In the extreme, interdisciplinary integrated analysis may lead to **transdisciplinary science**, that is, the modification of disciplinary, and the emergence of new interdisciplinary, concepts, methods or theories. They are firmly based on established concepts, methods or theories of one particular discipline, which are modified so as to fit with concepts, methods or theories from another discipline. If the original disciplines do not accept this modification, it may lead to the emergence of a new scientific discipline which is defined by its subject matter and its concepts, methods and theories. The reference to the disciplines from which it has emerged is then purely historical. An example is the field of molecular biology, which has emerged as an independent discipline from the integration of concepts, methods and theories from biology, chemistry and physics. But it is also possible that the interdisciplinary modifications act back on the original disciplines and leave a permanent impact on them.

(vi) The notion of **transdisciplinary problem solving** is sometimes used in an even wider meaning to denote cooperation beyond the boundaries of science, e.g. with stakeholders or practitioners disposing of non-scientific knowledge. The discourse with, and participation of, such social actors and groups should help to identify relevant research questions and conceptual structures of some problem under study, which later on facilitates the adoption and implementation of solutions. An example is the search for, and sustainable management of, pharmaceutical substances embedded in the naturally occurring biodiversity.[10] This endeavor brings together academic scientists – such as biologists, chemists and physiologists – and indigenous people with their traditional knowledge about the medicinal impact of local plants.

One cannot generally say that one of these approaches is superior to the others. All of them have merit in some respect and shortcomings in some other respect. Which approach to follow when combining insights from different disciplines depends on the scientific aims and the subject matter to be studied.

For the purpose of ecological, environmental and resource economics, and the purpose of this study, approach (i) will clearly not do because it cannot guarantee that exactly those natural science insights which are relevant for economics are taken up. This would require at least some minimal exchange with economics and, thus, go beyond a multidisciplinary analysis. Also, a multidisciplinary approach cannot guarantee that natural science insights are put forward in a manner compatible with the conceptual structure and terminology of economics. This requires an interdisciplinary division of labor or integrated analysis of economics and the natural sciences. On the other hand, the approaches (v) and (vi) are very ambitious and go far beyond standard science. So, the analysis in this study is interdisciplinary in the sense of approaches (iii) and (iv). While the procedure in Part I (Thermodynamics) is predominantly characterized by an interdisciplinary division of labor between economics and thermodynamics, the procedure in Part II (Biodiversity) is predominantly characterized by an interdisciplinary integrated analysis between economics and ecology. This will be explained in more detail in the following.

[10]See the detailed discussion of this approach to 'bio-prospecting' on page 127.

1.3.3 Method Orientation and Problem Orientation

Any scientific analysis studies a certain subject matter (*problem*) with a certain toolbox of concepts, methods and theories (*method*). Accordingly, any scientific research program may be driven either by the primary purpose of better understanding a certain subject matter (*solving a problem*) or by the primary purpose of further elaborating and advancing the methodological toolbox of science (*enhancing a method*).[11] The same goes for interdisciplinary science: the purpose of interdisciplinary science, or the challenge arising from interdisciplinary cooperation, may be primarily related either to solving a particular problem or to enhancing a particular scientific method.

In this study, I follow both approaches to the interdisciplinary integration of natural science constraints into ecological, environmental and resource economics: Part I of this study is primarily motivated by enhancing a particular method – thermodynamic analysis in ecological, environmental and resource economics; Part II is primarily motivated by contributing to the solution of a particular problem – biodiversity loss and conservation. Both the particular method and the particular problem under study are typical examples for the respective approach. In each approach, the interdisciplinary challenge of how to integrate natural science constraints into ecological, environmental and resource economics has a different character and, therefore, entails a different kind of solution. Hence, Parts I II of this study yield complementary insights into how to conceptualize natural science constraints in ecological, environmental and resource economics.

The method-oriented approach

In the method-oriented approach of Part I of this study, the method of thermodynamics serves to conceptualize constraints on economic action, in particular on production processes which can be described as a transformation of energy and matter. The two most elementary constraints stem from the first and second laws of thermodynamics: conservation of mass (First Law) and irreversibility (Second Law). These constraints are exogenous to economic action and fixed. Therefore, in order to capture thermodynamic constraints in ecological, environmental and resource economics, an interdisciplinary division of labor between thermodynamics and economics is possible. The task of thermodynamics in this division of labor is to conceptualize and formalize these constraints so that they are compatible with the conceptual structure and terminology of economics. For instance, the laws of thermodynamics can impose restrictions on the set of feasible economic allocations. Once this first step has been completed, it is the task of economics to then study which alloca-

[11] Of course, the motivation of a scientific research program may also be a combination of these two polar cases.

tions are, and should be, chosen individually and collectively. Since this second step of the analysis does not have any repercussions on the first step, a clear division of labor between thermodynamics and economics is possible. As a result, the method-oriented approach is much closer to the categorically distinct approaches of the two disciplines of thermodynamics and economics than the problem-oriented approach.

The problem-oriented approach

Part II of this study follows a problem-oriented approach to studying biodiversity loss and conservation. This is a complex and multifarious real-world problem at the intersection of economies and ecosystems. Being a real-world phenomenon, it first has to be translated into scientific terms before it can be studied by scientific means. This is a challenge that does not occur in the method-oriented approach, as the starting point of the method-oriented approach is already – by definition – within the realm of science. Any real-world problem can be translated such that it falls into the domain of one discipline or the other, or a certain set of disciplines. While there is considerable freedom (even arbitrariness) as to how to translate a real-world phenomenon into scientific terms, it then is up to the different disciplines involved in an interdisciplinary problem-oriented analysis to make sure that their respective contributions really deal with the same phenomenon and fit with each other. In most cases, an interdisciplinary division of labor will not do for that purpose. Rather, a problem-oriented approach most often requires an interdisciplinary integrated analysis.

In the case of the problem-oriented analysis of biodiversity loss and conservation, there is another reason why an interdisciplinary division of labor will not do, but an interdisciplinary integrated analysis is required. As the problem is at the intersection of two mutually interacting systems, the ecological system and the economic system, there are many potential feedbacks from one system onto the other. Due to these feedbacks, a clear division of labor between the natural sciences and economics is not possible, because in a division of labor one would loose the feedbacks from the analysis. Instead, an encompassing analysis of complex problems, such as biodiversity loss and conservation, requires an interdisciplinary integrated analysis (Wätzold et al., forthcoming).

The natural science constraints, which in the case of biodiversity mainly come from ecology, can no longer be taken to be fixed and exogenous to ecological, environmental and resource economics. Instead, while ecology still imposes constraints on economic action, the choice of a particular allocation by an economic agent has repercussions on the ecological system, which, in turn, influences the ecological constraints. These feedbacks have to be taken into account in a problem-oriented ecological-economic analysis of biodiversity loss and conservation. As a consequence, the analysis of biodiversity loss and conservation in Part II of this study is an interdisciplinary integrated analysis, in which the economic system is modelled based on economic concepts and relationships, and the ecological system is modelled based on ecological concepts and relationships, with both sets of concepts and relationships highly adapted to each other and to the joint research aims, so that the impact of economic action onto the ecological system, and the resulting feedback for the set of feasible economic actions, can be studied explicitly.

1.3.4 Conceptual Analysis

This study is a conceptual (as opposed to: empirical or applied) analysis. It focuses on the definition, clarification and interdisciplinary application of *concepts* that are of central relevance for structuring the analysis of economy-environment interactions from the interdisciplinary perspective of economics and the natural sciences. This includes the exploration of relationships among these concepts with the help of conceptual models. It leads to the development of policy recommendations on a conceptual level, e.g. for the control of emissions (Chapter 6) or for ecosystem management (Chapter 11). Examples of concepts, which are central to this study, include conservation of mass, irreversibility and joint production in Part I; and biodiversity, ecosystem services and insurance in Part II.

Conceptual analysis is an accepted approach in modern economics, and has been advocated before on several occasions, also for the field of ecological, environmental and resource economics. For example, in their pioneering book on *Economic Theory and Exhaustible Resources*, Dasgupta and Heal (1979: 9–10) justify conceptual analysis as follows:

> What one aims at in constructing an economic model, whose purpose is the development of understanding at a basic conceptual level (as opposed for example to the prediction of the values to be assumed by a particular set of variables at a future date), is to strip away detail and in the process sacrifice precision, in order to grasp at general principles which would be obscured but by no means invalidated by the inclusion of detail. What one aims at in other words is the construction of a framework which is simple enough to reveal the principles at work but whose basic structure is robust to the kinds of additions and extensions generally needed to implement the analysis in any particular situation.

Beyond its use in individual scientific disciplines, conceptual analysis is necessary and essential in order to lay a solid basis for the interdisciplinary integration of established scientific disciplines. It helps to identify the potential as well as the pitfalls of interdisciplinary integration and is a prerequisite for asking relevant and meaningful research questions (cf. Section 1.3.2).

After the methodological basis is now prepared, the remaining sections of this chapter introduce the contents of this study. Section 1.4 previews Part I (Thermodynamics), while Section 1.5 previews Part II (Biodiversity).

1.4 The Method of Thermodynamics

This section gives a brief outline of the analysis and results in **Part I** of this study (Chapters 2–6), which deals with thermodynamic analysis of economy-environment interactions and is characterized by method-orientation. Thermodynamics is the branch of physics that deals with macroscopic transformations of energy and matter. The origins of thermodynamics are to be found in the nineteenth century when practitioners, engineers and scientists like James Watt (1736–1819), Sadi Carnot (1796–1832), James Prescott Joule (1818–1889), Rudolph Clausius (1822–1888) and William Thomson (the later Lord Kelvin, 1824–1907) wanted to understand and increase the efficiency with which steam engines perform useful mechanical work. From the beginning, this endeavor has combined the study of natural systems and the study of engineered systems – created and managed by purposeful human action – in a very peculiar way, which is rather unusual for a traditional natural science such as physics.

Not surprisingly then, the laws of thermodynamics were found by economists to be concepts with considerable implications for economics. In the late 1960s and early 1970s economists discovered the relevance of thermodynamics for environmental and resource economics (Pethig 2003, Spash 1999: 418, Turner 1999a: Section 2). For instance, economists like Kenneth Boulding (1966), Robert Ayres and Allen Kneese (1969), and Nicolas Georgescu-Roegen (1971) turned to thermodynamics when they wanted to analyze economy-environment interactions in an encompassing way, and root the economy in its biogeophysical basis analytically.

1.4.1 Thermodynamic Analysis: Rationale, Concepts, and Caveats

Chapter 2 opens the discussion of how to integrate thermodynamic analysis into ecological, environmental and resource economics. It lays out the fundamental rationale of this endeavor, which is crucially based on the duality between real and monetary descriptions of economic action, and sketches its historical origins. It also addresses the question 'How can thermodynamic concepts, laws and results be incorporated in a fruitful manner into economic analysis?' This has been attempted in four basic ways, which are very different in the intellectual approach taken:

1. Isomorphism of formal structure,

2. Analogies and metaphors,
3. Energy, entropy and exergy theories of value,
4. Thermodynamic constraints on economic action.

It is the last one of these approaches, which is taken in this study. It builds on a clear division of labor between the disciplines of thermodynamics and economics. The laws of thermodynamics are used to capture the constraints on transformations of energy and matter. Their role is limited to this particular task. Based on this conceptualization of constraints, methods and concepts from economics are then used to study allocations in an economy which result from the optimizing behavior of firms and households, e.g. profit-maximizing resource-extraction and production firms as well as utility-maximizing households purchasing the consumer goods produced. This approach can be operationalized directly, and is empirically meaningful for ecological, environmental and resource economics. It lends itself quite naturally to modeling.

Chapter 2 also surveys the literature on implications and insights of thermodynamic analysis in ecological, environmental and resource economics. The chapter concludes by assessing the role of thermodynamics for ecological, environmental and resource economics, and for the discussion of sustainability. There is a brief and basic introduction into the elementary concepts and laws of thermodynamics in the Appendix.

1.4.2 The Inada Conditions for Material Resource Inputs Reconsidered

Chapter 3 formally explores one particular implication that the thermodynamic law of conservation of mass, the so-called Materials-Balance-Principle, has for modeling production. It is shown that the marginal product as well as the average product of a material resource input are bounded from above. This means that the usual Inada conditions (Inada 1963), when applied to material resource inputs, are inconsistent with a basic law of nature. This is important since the Inada conditions are usually held to be crucial for establishing steady state growth under scarce exhaustible resources.

While the advocates of a thermodynamic-limits-to-economic-growth perspective (e.g. Boulding 1966, Daly 1991[1977], Georgescu-Roegen 1971) usually stress the universal and inescapable nature of limits imposed by laws of nature, pro-economic-growth advocates usually claim that there is plenty of scope for getting around particular thermodynamic limits by substitution, technical progress and 'dematerialization' (e.g. Beckerman 1999, Smulders 1999, Stiglitz 1997). The latter therefore often conclude that, on the whole, thermodynamic constraints are simply irrelevant for economics. This chapter takes a more differentiated stand, by analyzing in detail

(i) what exactly are the implications of thermodynamics for modeling production at the level of a single production process, and

(ii) how these constraints carry over to the level of aggregate production, considering that there is scope for substitution in an economy between different resources and different production technologies.

1.4.3 Temporal and Thermodynamic Irreversibility in Production Theory

From a physical point of view, irreversibility is an essential dynamic feature of real production. Therefore, it should be properly taken into account in dynamic analyses of production systems. The idea of irreversibility can be rigorously rooted in the laws of thermodynamics (Kondepudi and Prigogine 1998: 84ff, Zeh 2001). The importance of thermodynamic irreversibility, and the physicists' preoccupation with this concept, lies in the fact that it precludes the existence of perpetual motion machines, that is, devices which use a limited reservoir of available energy to perform work forever (Second Law of Thermodynamics). It is an everyday experience that no such thing as a perpetual motion machine exists. In order to make this insight accessible to economic analysis, and to the study of long term economy-environment interactions, it is necessary to adequately represent thermodynamic irreversibility as a constraint for economic action (Georgescu-Roegen 1971).

Economists have devoted some effort to incorporating irreversibility into production theory. However, irreversibility has often been introduced into the theory as an ad-hoc-assumption. As a result, the assumption did not always achieve what it actually should achieve from a thermodynamic point of view, namely to imply irreversibility of the system's evolution as stated by the Second Law of Thermodynamics.

Chapter 4 introduces a formal and rigorous definition of thermodynamic irreversibility, which is (i) sound from a physical point of view and (ii) formulated such that it is compatible with formal modelling in economic production theory. In order to assess, whether – and to what extent – different notions of irreversibility from production theory capture thermodynamic irreversibility, two prominent irreversibility concepts – the one due to Koopmans (1951b) and the one due to Arrow-Debreu (Arrow and Debreu 1954, Debreu 1959) – are reexamined against the definition of thermodynamic irreversibility. It is shown that Koopmans' notion of irreversibility fully captures thermodynamic irreversibility, and that the notion of Arrow-Debreu, which has become the standard one in economic theory, does not capture thermodynamic irreversibility but only the weaker aspect of temporal irreversibility. This means, the standard irreversibility concept of production theory is too weak to be in full accordance with the laws of nature.

1.4.4 Necessity and Inefficiency in the Generation of Waste

It has been argued, based on the thermodynamic laws of mass conservation and entropy generation, that in industrial production processes the occurrence of waste is as necessary as the use of material resources (Ayres and Kneese 1969, Faber et al. 1998, Georgescu-Roegen 1971).[12] On the other hand, it seems to be quite obvious that the sheer amount of waste currently generated in modern industrial economies is to some extent due to various inefficiencies and might, in principle, be reduced.

Chapter 5 discusses to which extent the occurrence of waste is actually an unavoidable necessity of industrial production, and to which extent it is an inefficiency that may, in principle, be reduced. For that sake, the laws of thermodynamics are employed as an analytical framework within which results about current 'industrial metabolism' (Ayres and Simonis 1994) may be rigorously deduced in energetic and material terms. It is demonstrated that the occurrence of waste by-products is an unavoidable necessity in the industrial production of desired goods. While waste is thus an essential qualitative element of industrial production, the quantitative extent to which waste occurs may vary within certain limits according to the degree of thermodynamic (in)efficiency with which these processes are operated. The chapter discusses the question of which proportion of the amount of waste currently generated is due to thermodynamic necessity, and which proportion is due to thermodynamic inefficiency.

1.4.5 Optimal Dynamic Scale and Structure of a Multi-Pollution Economy

While Chapters 3–5 have dealt with the implications from thermodynamics for *modelling* the production process at both the micro- and macro-level, the question of *optimal allocation* of resources has not been addressed so far. **Chapter 6** closes this gap. It takes as its starting point a thermodynamic representation of production as joint production of consumption and environmental pollution, and explores the implications for optimal macroeconomic dynamics.

Chapter 6 looks into the coupled environmental-economic dynamics of a multi-sector-multi-pollution-economy. It addresses the following questions: How should the macroeconomic scale and structure change over time in response to the dynamics of environmental pollution?[13] Is this dynamic process monotonic over time, or can a trade-off between long-run and short-run considerations

[12]For example, Georgescu-Roegen (1975: 357) has argued that 'waste is an output just as unavoidable as the use of natural resources'.

[13]*Scale* means the overall level of economic activity, measured by total factor input; and *structure* means the composition of economic activity, measured by relative factor inputs to different sectors.

(e.g. lifetime versus harmfulness of pollutants) induce a non-monotonic economic dynamics? What is the time scale of economic dynamics (i.e. change of scale and structure), and how is it influenced by the different time scales and constraints of the economic and environmental systems? These questions are relevant for the current policy discussion on the sustainable biophysical scale of the aggregate economy relative to the surrounding natural environment (e.g. Arrow et al. 1995, Daly 1992a, 1996, 1999), and how economic policy should promote structural economic change as a response to changing environmental pressures (e.g. de Bruyn 1997, Winkler 2005).

The analysis shows that along the optimal time-path (i) the overall scale of economic activity may be less than maximal, (ii) the time scale of economic dynamics (change of scale and structure) is mainly determined by the longest-lived pollutant, (iii) the optimal control of emissions may be non-monotonic. In particular the last result raises important questions about the design of optimal environmental policies.

1.5 The Problem of Biodiversity Loss and Conservation

This section gives a brief outline of the analysis and results in **Part II** of this study (Chapters 7–11), which deals with biodiversity loss and conservation and is characterized by problem-orientation. Biological diversity (or 'biodiversity', for short), which has been defined as 'the variability among living organisms from all sources ... and the ecological complexes of which they are part' (CBD 1992), is valuable for humans for a number of reasons. Many species have direct use value as food, fuel, construction material, industrial resource or pharmaceutical substance. Biodiversity also has an important indirect use value in so far as entire ecosystems perform valuable services such as nutrient cycling, control of water runoff, purification of air and water, soil regeneration, pollination of crops and natural vegetation, control of pests and diseases, or local climate stabilization (Daily 1997b, Millennium Ecosystem Assessment 2005). These ecosystem services can only be provided by more or less intact ecosystems and result from the complex – and up to now not well understood – interplay of many different species in these ecosystems (Holling et al. 1995, Hooper et al. 2005, Kinzig et al. 2002, Loreau et al. 2001, 2002b, Schulze and Mooney 1993, Tilman 1997a).

Biodiversity is currently being lost at rates that exceed the natural extinction rates of the past by a factor of somewhere between 100 and 1,000 (Watson et al. 1995b). This is one of the most eminent environmental problems of our time (Wilson 1988). By now, the international community has acknowledged the problem of biodiversity loss, and the need to enact policies to halt or even reverse this problem. For example, in June 1992, the *Convention on Biological Diversity* was signed by 156 states at the United Nations Conference on Environment and Development in Rio de Janeiro, Brazil, with the aim of safeguarding the sustainable conservation and use of biodiversity at the global level (CBD 1992).

1.5.1 Biodiversity as an Economic Good

Chapter 7 opens the discussion of biodiversity loss and conservation by addressing, on a fundamental level, the question of what economics can contribute to an encompassing discussion of biodiversity loss and conservation. More specifically, it addresses the following questions:

(i) In what sense can one think of biodiversity as an economic good?

(ii) In what sense does biodiversity have economic value?

(iii) What can economic analysis contribute to the explanation of biodiversity loss?

(iv) What is the relevance of economic valuation for biodiversity conservation?

Discussing these questions helps to clarify the conceptual foundations upon which an ecological-economic analysis of biodiversity loss and conservation is possible. At the same time, it sheds light on the question of how exactly the two disciplines of economics and ecology need to interact in order to generate fruitful and relevant contributions to this analysis.

The chapter is written from an economic perspective and serves as a survey of the relevant literature. As its starting point, it takes the hypothesis that biological diversity can be thought of as an economic good which has economic value. This hypothesis is vindicated in detail. Its fruitfulness is then tested by applying it to explain the large-scale loss of biodiversity currently observed and to develop recommendations for biodiversity conservation. Viewing biodiversity as an economic good, which has economic value, makes obvious the potential and limits of economics as an academic discipline for the discussion of biodiversity loss and conservation. These insights form the working basis of the remaining chapters of Part II, which display different degrees of interdisciplinary integration of economics with ecology.

1.5.2 Ecological and Economic Measures of Biodiversity

For analyses of how biodiversity contributes to ecosystem functioning, how it enhances human well-being, and how these services are currently being lost, a quantitative measurement of biodiversity is crucial. Ecologists, for that sake, have traditionally employed different concepts such as species richness, Shannon-Wiener-entropy, or Simpson's index (see e.g. Magurran 2004). Recently, economists have added to that list measures of (bio)diversity that are based on pairwise dissimilarity between species (e.g. Weitzman 1992, 1998) or, more generally, weighted features of species (Nehring and Puppe 2002, 2004).

In **Chapter 8**, I give a review and conceptual comparison of the two broad classes of biodiversity measures currently used, the ecological ones and the economic ones. It turns out that the two classes are distinct by the information they use for constructing a diversity index. While the ecological measures use the number of different species in a system as well as their relative abundances, the economic ones use the number of different species as well as their characteristic features. In doing so, the two types of measures aim at characterizing two very different aspects of the ecological-economic system. The economic measures characterize the abstract list of species existent in the system, while the ecological measures target the actual, and potentially unevenly distributed allocation of species.

I argue that the underlying reason for this difference is in the philosophically distinct perspective on diversity between ecologists and economists. Ecologists traditionally view diversity more or less in what may be called a 'conservative' perspective, while economists predominantly adopt what may be called a 'liberal' perspective on diversity (Kirchhoff and Trepl 2001). In the conservative view, which goes back to Leibniz and Kant, diversity is an expression of unity. By viewing a system as diverse, one stresses the integrity and functioning of the entire system. The ultimate concern is with the system at large. In this view, diversity may have an indirect value in that it contributes to certain overall system properties, such as stability, productivity or resilience at the system level. In contrast, in the liberal view, which goes back to Descartes, Hume and Locke, diversity guarantees the freedom of choice for autonomous individuals who choose from a set of diverse alternatives. The ultimate concern is with the well-being of individuals. In this view, diversity of a choice set has a direct value in that it allows individuals to make a choice that better satisfies their individual and subjective preferences.

The question of how to measure biodiversity, thus, is ultimately linked to the question of what is biodiversity good for. Do we see biodiversity as valuable for individuals who want to make a choice from a diverse resource base, e.g. when choosing certain desired genetic properties of crops or pharmaceutical substances? Or do we see it as valuable for overall ecosystem functioning, e.g. out of a concern for conserving certain desired ecosystem services such as water purification or soil regeneration? Of course, there is a continuous spectrum in between these two extremes. But in any case, so the conclusion of this chapter, the measurement of biodiversity requires a prior normative judgment as to what purpose biodiversity serves in ecosystems and economic systems.

1.5.3 The Insurance Value of Biodiversity in the Provision of Ecosystem Services

Biodiversity is useful and valuable to humans for a number of reasons (see the discussion in Chapter 7). One particular reason is that biodiversity provides insurance by stabilizing the provision of ecosystem services which are being used by risk-averse economic agents. In **Chapter 9**, I present a conceptual ecological-economic model that combines (i) ecological results about the relationships between biodiversity, ecosystem functioning, and the provision of ecosystem services with (ii) economic methods to study decision-making under uncertainty. In this framework I (1) determine the insurance value of biodiversity, (2) study the optimal allocation of funds in the trade-off between investing into biodiversity protection and the purchase of financial insurance, and (3) analyze the effect of different institutional settings in the market for financial insurance on biodiversity protection.

The conclusion from this analysis is that biodiversity can be interpreted as a form of natural insurance for risk averse ecosystem managers against the over- or under-provision with ecosystem services, such as biomass production, control of water run-off, pollination, control of pests and diseases, nitrogen fixation, soil regeneration etc. Thus, biodiversity has an insurance value, which is a value component in addition to the usual value arguments (such as direct or indirect use or non-use values, or existence values) holding in a world of certainty. This insurance value should be taken into account when deciding upon how much to invest into biodiversity protection. It leads to choosing a higher level of biodiversity than without taking the insurance value into account, with a higher degree of risk aversion leading to a higher optimal level of biodiversity. As far as the insurance function is concerned, biodiversity and financial insurance against income risk, e.g. crop yield insurance, may be seen as substitutes. If financial insurance is available, a risk averse ecosystem manager, say, a farmer, will partially or fully substitute biodiversity's insurance function by financial insurance, with the extent of substitution depending on the costs of financial insurance. Hence, the availability, and exact institutional design, of financial insurance influence the level of biodiversity protection.

1.5.4 Insurance and Sustainability through Ecosystem Management

As shown in Chapter 9, biodiversity has an insurance value which is relevant for decisions about how to manage ecosystems. While the analysis in Chapter 9 was based on a very simple and stylized ecological-economic model, in order to focus on the conceptual structure of the argument, **Chapter 10** develops this argument further by looking in detail at a realistic case: grazing management in semi-arid rangelands. Livestock farmers in semi-arid regions make use of the ecosystems' insurance function by choosing grazing management strategies so as to hedge against their income risk which stems from the stochasticity of precipitation.

The analysis in Chapter 10 is based on a dynamic and stochastic ecological-economic model of grazing management in semi-arid rangelands. The non-equilibrium ecosystem is driven by stochastic precipitation. A risk averse farmer chooses a grazing management strategy under uncertainty so as to maximize expected utility from farming income. Grazing management strategies are rules about which share of the rangeland is given rest depending on the actual rainfall in that year. In a first step, the farmer's short-term optimal grazing management strategy is determined. It is shown that a risk-averse farmer chooses a strategy so as to obtain insurance from the ecosystem: the optimal strategy reduces income variability, but yields less mean income than possible. In a second step, the long-run ecological and economic impact of different strategies is analyzed. The conclusion is that the more risk-averse a farmer is, the more conservative and sustainable is his short-term optimal grazing management strategy, even if he has no specific preference for the distant future.

1.5.5 Optimal Investment in Multi-Species Protection

From the discussion in Chapter 7 it has become apparent that biodiversity is useful and valuable to humans for many reasons, with one particular reason – its insurance function – discussed in detail in Chapters 9 and 10. It has also become apparent that biodiversity is currently being lost at, on average, suboptimally high rates. This raises the question of how to protect biodiversity in a manner that is ecologically effective and economically efficient.

In **Chapter 11**, I contribute to this discussion by studying optimal investment in multi-species protection when species interact in an ecosystem. The analysis is based on a model of stochastic species extinction in which survival probabilities are interdependent. Individual species protection plans can increase a species' survival probability within certain limits and contingent upon the existence or absence of other species. Protection plans are costly and the conservation budget is fixed. It is assumed that human well-being depends solely on the services provided by one particular species, but other species contribute to overall ecosystem functioning and thus influence the first species' survival probability.

The analysis shows that taking into account species interactions in an ecosystem is crucial for the optimal allocation of a conservation budget. Compared with policy recommendations obtained under the assumption of independent species, interactions in an ecosystem can reverse the rank ordering of spending priorities among species conservation projects. Hence, an approach to species protection that is efficient in terms of both species conservation and budget resources should be based on a multi-species framework and should take into account the basic underlying ecological relations. Another interesting result is that even if biological conservation decisions are exclusively derived from a utilitarian framework, with species interaction it may be optimal to invest in the protection of species that do not directly contribute to human well-being. This is due to their role for overall ecosystem functioning and for safeguarding the existence of those species that are the ultimate target of environmental policy. The conclusion is that effective species protection should go beyond targeting individual species, and consider species relations within whole ecosystems as well as overall ecosystem functioning.

PART I

Thermodynamics

2. Thermodynamic Analysis in Ecological, Environmental and Resource Economics: Rationale, Concepts, and Caveats*

2.1 Introduction

Integrating methods and models from thermodynamics and from economics promises to yield encompassing insights into the nature of economy-environment interactions. At first sight, the division of labor between thermodynamics and economics seems obvious. Thermodynamics should provide a description of societies' physical environment, while economics should provide an analysis of optimal individual and social choice under the restriction of environmental scarcities.

But the task is more difficult. Being a branch of physics, thermodynamics is a natural science. It explains the world in a descriptive and causal, allegedly value-free manner. On the other hand, economics is a social science. While it pursues descriptive and causal (so-called 'positive') explanations of social systems to a large extent, it also has a considerable normative dimension. Valuation is one of its basic premises and purposes. Bringing together thermodynamics and economics in a common analytical framework therefore raises all kinds of questions, difficulties and pitfalls.

This chapter discusses the rationale, concepts, and caveats for integrating thermodynamic analysis into ecological, environmental and resource economics. Section 2.2 lays out the fundamental rationale of this endeavor and sketches its historical origins. Section 2.3 identifies different approaches to incorporating thermodynamic concepts into economic analysis and assesses their respective potential for ecological, environmental and resource economics. Section 2.4 surveys various implications and insights that thermodynamic analysis has already yielded for ecological, environmental and resource economics. Section 2.5

*Revised version of 'Thermodynamic Models', previously published in J. Proops and P. Safonov (eds), *Modelling in Ecological Economics*, Cheltenham: Edward Elgar, 2004, pp. 102–129.

concludes by assessing the role of thermodynamics for ecological, environmental and resource economics, and for the discussion of sustainability. There is a brief and basic introduction into the elementary concepts and laws of thermodynamics in the Appendix.

2.2 Fundamental Rationale and Historical Origins

2.2.1 Different Perspectives on Economy-Environment Interactions

When economists started to analyze the flow of resources, goods, services and money in an economy, the picture was pretty simple: there are two groups of economic agents, consumers and producers; producers deliver goods and services to consumers, and consumers provide the resources with which they are endowed, labor in particular, to producers. Thus, there is a circular flow of commodities in an economy. There is an equivalent circular flow of money counter to that primary flow, as consumers pay money to producers for the goods they consume, and producers remunerate the labor force they receive from the consumers/laborers.[1]

Since the two corresponding flows, the primal flow of real commodities and the dual flow of monetary compensation, are exactly equivalent, it seems superfluous to always study both of them when analyzing economic transactions and allocations. Hence, the convention was established in economics to focus on the monetary flow. The current system of national economic accounts, which is meant to be a full representation of economic activity in an economy over one time period, therefore captures all transactions in monetary units, e.g. the provision of labor and capital, the trading of intermediate goods and services between different sectors of the economy, and final demand for consumer goods.

Of course, this picture is too simple. It neglects the use of natural resources and the emission of pollutants and wastes. Both activities are unavoidable aspects of economic action (see Chapter 5 below). In the early twentieth century, the subdiscipline of *environmental and resource economics* emerged to deal with the question of how to account, in an economic sense, for the use of natural resources on the one hand and the emission of pollutants and wastes on the other (Gray 1913, 1914, Hotelling 1931, Pigou 1912, 1920). The picture now appeared as follows: there is a circular flow – actually: two equivalent circular flows – between consumers and producers which form the core of economic activity. In addition, there is an inflow of natural resources and an outflow of

[1]Later, this system was extended to include savings and investment, as well as imports and exports.

emissions and wastes. Thus, a linear throughflow of energy and matter drives the circular flow of economic exchange.

A further step in the development of thinking about economy-environment interactions, was the insight that the inflow of natural resources (resource economics) and the outflow of emissions and wastes (environmental economics) are not independent. Obviously, these two flows are linked by economic activity, i.e. economic activity transforms natural resources into emissions and wastes. But these two flows are also linked because they originate and terminate in the natural geobiophysical environment. For example, environmental pollutants released into natural ecosystems may impair the ecosystems' ability to produce the ecosystem goods and services, which are then used as a natural resource by the economy. This means, the extraction of natural resources, the production of goods and services within the economy, as well as the emission of pollutants and wastes all happen within the system of the natural geobiophysical environment.

This is the 'vision' (in the sense of Schumpeter)[2] of *ecological economics*: ecological economics views the human economy as an open subsystem of the larger, but finite, closed, and non-growing system of non-human nature (Ayres 1978, Boulding 1966, Daly 1991[1977], Faber and Proops 1998, Georgescu-Roegen 1971, and many more). In this view, the human economy is a part of nature. In contrast, in the view of traditional environmental and resource economics Nature is treated as a part of the human economy. Both 'resources' and 'environment' are treated as additional economic sectors in the system of national economic accounts, and flows to and from these sectors are accounted for in monetary units.[3]

2.2.2 Duality Between the Real and Monetary Descriptions and the Role of Thermodynamics

Environmental and resource economics faced one conceptual problem from the very beginning. Economic analysis, including environmental and resource economics, is based on the idea of duality (i.e. equivalence) between the flow of real commodities and services (measured in physical units) and an equivalent value flow (measured in monetary units), and consequently focuses on the value dimension. But the inflow of natural resources, as well as the outflow of emissions and waste, do not have an apparent value dimension. Markets do not indicate these values, as markets often do not exist in this domain. And where

[2]Schumpeter (1954: 42) defines a *vision* as the 'preanalytic cognitive act that necessarily precedes any scientific analysis'.

[3]Brown (2001: 5) argues that the change of perspective from nature-as-part-of-the-economy to the-economy-as-part-of-nature amounts to a scientific revolution not unlike the transition from the geocentric to the heliocentric world view in the Copernicanean revolution.

they exist, the resulting values are distorted due to ubiquitous externalities and public goods.

As a result, the valuation of natural goods and services has to be set up explicitly as a non-market process, and elaborate theories and techniques have been proposed for this purpose.[4] All these techniques require, to a greater or lesser extent, an adequate, prior description – in real terms – of the particular commodity or service to be valued. In other words, before individuals or society can value something, they have to have an adequate idea about what exactly that something is. This holds, in particular, for the energy and material resources used in production as well as for the emissions and wastes generated as by-products of desired goods.

And here lies the relevance of thermodynamics. Being the branch of physics that deals with transformations of energy and matter, thermodynamics is an appropriate foundation in the natural sciences to provide a description in real terms of what goes on when humans interact with the non-human environment. In particular, thermodynamics captures the energy/matter dimension of economy-environment interactions. Thus, it is a necessary complement and prerequisite for economic valuation.

2.2.3 Historical Origins of Thermodynamic Analysis in Ecological, Environmental and Resource Economics

The origins of thermodynamics are to be found in the nineteenth century when practitioners, engineers and scientists like James Watt (1736–1819), Sadi Carnot (1796–1832), James Prescott Joule (1818–1889), Rudolph Clausius (1822–1888) and William Thomson (the later Lord Kelvin, 1824–1907) wanted to understand and increase the efficiency with which steam engines perform useful mechanical work. From the beginning, this endeavor has combined the study of natural systems and the study of engineered systems – created and managed by purposeful human action – in a very peculiar way, which is rather unusual for a traditional natural science such as physics.

Not surprisingly then, the laws of thermodynamics were found by economists to be concepts with considerable implications for economics. In the late 1960s and early 1970s economists discovered the relevance of thermodynamics for environmental and resource economics (Pethig 2003, Spash 1999: 418, Turner 1999a: Section 2). For instance, economists like Kenneth Boulding (1966), Robert Ayres and Allen Kneese (1969), and Nicolas Georgescu-Roegen (1971) turned to thermodynamics when they wanted to analyze economy-environment interactions in an encompassing way, and root the economy in its biogeophysical basis analytically.

[4]For an overview see e.g. Freeman (2003) or Hanley and Spash (1993).

In a first step, the Materials Balance Principle was formulated based on the thermodynamic Law of Conservation of Mass (Ayres and Kneese 1969, Boulding 1966, Kneese et al. 1972). In view of this principle, all resource inputs that enter a production process eventually become waste. This is now an accepted and undisputed piece of ecological, environmental and resource economics.

At the same time, Georgescu-Roegen (1971) developed an elaborate and extensive critique of neoclassical economics based on the laws of thermodynamics, and, in particular, the Entropy Law, which he considered to be 'the most economic of all physical laws' (Georgescu-Roegen 1971: 280).[5] His contribution initiated a heated debate over the question of whether the Entropy Law – and thermodynamics in general – is relevant to economics (Burness et al. 1980, Daly 1992b, Kåberger and Månsson 2001, Khalil 1990, Lozada 1991, 1995, Norgaard 1986, Townsend 1992, Williamson 1993, Young 1991, 1994).[6] While Georgescu-Roegen had, among many other points, formulated an essentially correct insight into the irreversible nature of transformations of energy and matter in economies, his analysis is flawed to some extent by positing of what he calls a 'Fourth Law of Thermodynamics' (Ayres 1999b).[7] This may be the reason that the Second Law and the entropy concept have not yet acquired the same undisputed and foundational status for ecological, environmental and resource economics as have the First Law and the Materials Balance Principle.

But as Georgescu-Roegen's work and the many studies following his lead have shown, the Entropy Law, properly applied, yields insights into the irreversible nature of economy-environment interactions that are not available otherwise (Baumgärtner et al. 1996). Both the First and the Second Laws of Thermodynamics therefore need to be combined in the study of how natural resources are extracted, used in production, and give rise to emissions and waste, thus leading to integrated models of economy-environment interactions (e.g. Baumgärtner 2000, Faber et al. 1995, Perrings 1987, Ruth 1993, 1999).

2.3 Different Approaches

How can thermodynamic concepts, laws and results be incorporated in a fruitful manner into economic analysis? This has been attempted in four basic

---

[5] The works of Georgescu-Roegen are surveyed in a number of recent volumes (e.g. Beard and Lozada 1999, Mayumi 2001, Mayumi and Goody 1999) and a special edition of the journal *Ecological Economics* (Vol. 22, No. 3, 1997).

[6] See Baumgärtner et al. (1996) for a summary of this discussion.

[7] Georgescu-Roegen posited that in a closed system, matter is distributed in a more and more disordered way. He called this the 'Fourth Law', in extension of the three, well established laws of classical thermodynamics, described in the Appendix.

ways,[8] which are very different in the intellectual approach taken. In the following, I describe each of them in detail and assess their potential for ecological, environmental and resource economics.

2.3.1 Isomorphism of Formal Structure

Both thermodynamics and economics can be set up formally as problems of optimization under constraints. For example, equilibrium allocations in an economy can be viewed as a result of the simultaneous utility maximization under budget constraints of many households and profit maximization under technological constraints of many firms. Likewise, equilibrium micro- or macrostates of a thermodynamic system can be derived from the minimization of a thermodynamic potential, such as e.g. Helmholtz or Gibbs free energy, under the constraints of constant pressure, volume, chemical potential etc.[9] The mathematical structure of both economic and thermodynamic problems is, thus, formally equivalent. There is an isomorphism between the two types of problems and their respective solutions.[10]

As a result, one may exploit this formal isomorphism to obtain insights into the structure of economic equilibrium allocations from studying the structural properties of thermodynamic equilibria. Of course, these insights pertain to the formal structure of equilibrium solutions only, and do not contain any substantive content about thermodynamics or economics in themselves. For instance, based on what is known as the *Le Chatelier Principle* in thermodynamics (Kondepudi and Prigogine 1998: 239–240), Samuelson (1947, 1960a, 1960b) established the method of *comparative statics* in economics. This method explains the changes in the equilibrium solution of a constrained maximization problem (economic or thermodynamic) when one of the constraints is marginally tightened or relaxed. It has proven to be a very powerful tool and has found widespread use in modern economics.

Another example is the formal isomorphism between entropy and utility (Candeal et al. 2001a, 2001b), which becomes apparent in a particular entropy representation whereby entropy is constructed as an order preserving function that satisfies a continuity property (Candeal et al. 2001a, Cooper 1967).

---

[8] Söllner (1997) makes a similar distinction.

[9] This becomes most apparent in the Tisza (1966)/Callen (1985)-axiomatization of thermodynamics (Smith and Foley 2004, Sousa and Domingo, forthcoming).

[10] Mirowski (1989) has argued that modern economics is essentially built after the logic and formal structure of classical mechanics, i.e. Lagrangian and Hamiltonian formalism. Some authors in the ecological-economics community have taken this observation as a starting point for a methodological critique of the conventional economic approach to studying economy-environment interactions and proposed that ecological economics should be inspired more by thermodynamics instead, in order to get hands on the fundamental irreversibility of economy-environment interactions (Amir 1995, 1998, Costanza et al. 1997a, Georgescu-Roegen 1971, Lozada 1995, Martinez-Alier 1997).

But overall, it seems as if the potential of exploiting the isomorphism of formal structures in thermodynamic and economic equilibria was fairly limited and is, by now, largely exhausted.

2.3.2 Analogies and Metaphors

A second approach takes thermodynamic concepts and transfers them into economic thinking as analogies and metaphors (Faber and Proops 1985, Proops 1985, 1987). For example, under this approach, 'order' and 'disorder' in an economy are interpreted as expressions of 'social entropy', or the economy is seen as a 'self-organizing dissipative system far from thermodynamic equilibrium'. Typically, no attempt is made under this approach to clearly define the various terms, such as 'order', 'entropy' or 'equilibrium', in either thermodynamic or economic terms. Instead, these terms are used to evoke certain associations with the reader.

To a reader who is well trained in both thermodynamics and economics, it remains unclear whether a term like e.g. 'equilibrium' refers to thermodynamic equilibrium (in the sense of a thermodynamic system being in a state of minimal thermodynamic potential, e.g. Helmholtz free energy) or to economic equilibrium (in the sense of an economy of households and firms being in a state of market equilibrium where demand equals supply). Certainly, using these terms in such a loose manner cannot have the status of making exact and deductive scientific statements about economic systems.

Despite these large unclarities, the analogies-and-metaphors-approach has merit as a heuristic, since it allows one to see economic phenomena in a new light. Thus, it generates new and potentially fruitful questions, rather than answering existing ones. In that sense, it is more a 'vision' in the sense of Schumpeter (1954: 42),[2] than a rigorous analytical approach.

2.3.3 Energy, Entropy and Exergy Theories of Value

It has been argued that economic values based on subjective individual preferences are to some extent arbitrary and might be misleading in achieving sustainable solutions for environmental problems. In contrast, the argument goes, sustainability requires the identification of the 'true' and 'objective' value of nature's goods and services, and of damages to these. Often, thermodynamic quantities are proposed to give such an 'objective' value rod, e.g. energy (Costanza 1981, Hannon 1973, 1979, Hannon et al. 1986, Odum 1971), (low) entropy[11] or exergy (Bejan et al. 1996. 407).[12] In all these cases, the argument is essentially as follows: Energy (or, alternatively: exergy, low entropy) is the only really scarce factor here on Planet Earth. It therefore measures the ultimate scarcity that we face in dealing with nature. As a result, the amount of energy (exergy, low entropy) contained in every good or service measures its 'true' scarcity, and should therefore be taken as its value. Decisions concerning sustainability, so the argument, must be based on such energy/entropy/exergy-values, as they represent the ultimate scarcities.

From an economic point of view, this argument is untenable. It is untenable for the very same reasons that, for instance, a labor theory of value as advocated by David Ricardo or Karl Marx is untenable, and any other single-factor-theory of value would be untenable, be that factor energy, labor, oxygen, or anything else. 'Value', as it is understood in economics, results from the interplay of human goals and ends on the one hand (e.g. profit maximization, utility maximization, or sustainability), and scarcity of means to achieve these ends on the other hand (e.g. natural resources, capital, labor, or time). The higher the goals and the scarcer the resources necessary to achieve them, the more valuable are these resources. There is an economic theorem which states that only under very limiting assumptions the value of a good or service is given by the total amount of a factor of production (e.g. energy or labor) which has been used, directly or indirectly, in producing it. This is the so-called non-substitution theorem, proven in 1951 independently by four masterminds of economics: Arrow (1951), Koopmans (1951a), Georgescu-Roegen (1951) and Samuelson (1951).[13] This theorem identifies the conditions, under which a single-factor-theory of value holds:

(A1) There is only one primary, i.e. non-producible, factor of production.

(A2) This factor is directly used in the production of every intermediate or final good or service.

(A3) All production processes are characterized by constant returns to scale, i.e. scaling the amounts of all inputs by a factor of $\lambda > 0$ also scales the amount of output produced by the same factor λ.

---

[11] Burness et al. (1980: 7) and Patterson (1998) wrongly claim that Georgescu-Roegen (1971: Chapter 5) proposes a (low) entropy theory of value. On the contrary, Georgescu-Roegen (1971: 282) explicitly warns against such an interpretation. Georgescu-Roegen (1979) also gives an explicit rebuttal of energy theories of value. See Baumgärtner et al. (1996: 123–125) for details.

[12] Patterson (1998) surveys different theories of value in ecological economics.

[13] Three of these – Arrow, Koopmans, and Samuelson – were awarded the Nobel Prize in Economics. Some claim that the fourth one – Georgescu-Roegen – would have deserved it as well.

(A4) There is no joint production, i.e. every process of production yields exactly one output.

These are very restrictive assumptions. Only if (A1)–(A4) are fulfilled does a single-factor-theory of value fully explain the value of goods and services. If one of them does not hold, a single-factor-theory of value cannot provide a satisfactory explanation of value.

As for energy/entropy/exergy as a factor of production, one may safely assume that (A2) is fulfilled, and one may concede that (A1) can be taken to be fulfilled as well.[14] But in general, (A3) is not fulfilled, as many technologies are characterized by either increasing or decreasing returns to scale. And thermodynamic considerations, to which we will turn in detail later, imply that every process of production is joint production, so that (A4) is violated. This means, while energy, entropy or exergy theories of value are conceivable in very restricted models (characterized by Conditions A1–A4) they must be refuted for real economy-environment systems. To be sure, while energy, entropy or exergy are important factors in explaining value, value is a complex and encompassing phenomenon, and thermodynamic quantities alone cannot provide a satisfactory explanation.

2.3.4 Thermodynamic Constraints on Economic Action

Another approach to integrate insights from thermodynamics into economics starts from the observation that the laws of thermodynamics constrain economic action. Thermodynamic laws specify what is possible and what is not possible in the transformation of energy and matter. Such transformations play an important role in any economy, for example in

- the extraction of natural resources from the geo-bio-chemical-physical environment,
- the use of these resources in the production of goods and services,
- the generation and emission of wastes and environmental pollutants as by-products of desired goods, and
- the recycling of wastes into secondary resources.

All these transformations of energy and matter are at the center of interest in the field of ecological, environmental and resource economics. Hence, the laws of thermodynamics play an important role in describing relevant constraints

[14] One may also consider space and time as primary production factors, as they surely enter every process of production in some sense. But then, energy is not the only primary factor any more.

and scarcities for the economic analysis of economy-environment interactions (Cleveland and Ruth 1997).

This approach builds on a clear division of labor between the disciplines of thermodynamics and economics. The laws of thermodynamics are used to capture the constraints on transformations of energy and matter. Their role is limited to this particular task. Based on this conceptualization of constraints, methods and concepts from economics are then used to study allocations in an economy which result from the optimizing behavior of firms and households, e.g. profit-maximizing resource-extraction and production firms as well as utility-maximizing households purchasing the consumer goods produced.

This approach can be operationalized directly, and is empirically meaningful for ecological, environmental and resource economics. It lends itself quite naturally to modeling. One can distinguish between different model types for integrated thermodynamic-economic analysis, according to the thermodynamic concepts and laws they incorporate:

- models incorporating mass and conservation of mass (First Law), either for one particular material (say, copper) or for a number of materials,
- models incorporating energy and conservation of energy (First Law), sometimes in variants such as emergy ('embodied energy'),
- models incorporating entropy and entropy generation (Second Law),
- models incorporating energy and entropy, sometimes in the form of exergy (First and Second Law),
- models incorporating mass, energy and entropy (First and Second Law).

Models based on the First Law are useful for studying the economic implications from the scarcities due to physical conservation of mass and energy in the throughflow of materials and energy through the economy. Models based on the Second Law are useful for studying the economic implications from the scarcities due to the temporal directedness of this throughflow and its qualitative degradation by dissipation of energy and dispersal of matter.

2.4 Implications of and Insights from Thermodynamic Models

The use of thermodynamic concepts, laws and models in ecological, environmental and resource economics is an ongoing endeavor. So far, it has revealed a number of significant implications and insights about different aspects of economy-environment interactions.[15]

2.4.1 Materials Balance: The 'Planet Earth'-Perspective

The Materials Balance Principle is based on the Law of Conservation of Mass as implied by the First Law of Thermodynamics (Ayres 1978, Ayres and Kneese 1969, Boulding 1966, Kneese et al. 1972).[16] Since mass cannot be created, but is conserved in all transformations, all material resource inputs that enter a production process (i) diminish the corresponding resource reservoir, and (ii) eventually become waste.

This principle has lead to a view of the Earth, including the human society, as a 'spaceship' (Boulding 1966), which is completely closed to the surrounding space in material terms. Thus, all material transformations on Earth should be managed in a self-reliant and sustainable way.

2.4.2 Irreversibility of (Micro- and Macro-)Economic Processes

All processes of macroscopic change are irreversible. Examples include natural processes, such as the growing and blooming of a flower, as well as technical processes, such as the burning of fossil fuels in combustion engines. The entropy concept and the Second Law of Thermodynamics have been coined such as to capture this fact of nature (Kondepudi and Prigogine 1998, Zeh 2001).

The relevance of thermodynamic irreversibility for economics lies in the fact that it precludes the existence of perpetual motion machines, i.e. devices which use a limited reservoir of available energy to perform work forever. It is an everyday experience that no such thing as a perpetual motion machine exists. This holds for the micro-level, i.e. individual production processes, as well as for the macro-level, i.e. the economy at large (Georgescu-Roegen 1971).

In order to make this insight accessible to economic analysis it is necessary to adequately represent thermodynamic irreversibility as a constraint for economic action. Modern economic theory has devoted some effort to incorporating irreversibility into production theory. However, the standard irreversibility concept of economics, which is due to Arrow and Debreu (1954) and Debreu (1959), does not encompass thermodynamic irreversibility; it only establishes temporal irreversibility – a weaker form of irreversibility (Baumgärtner 2005).

---

[15] Surveys of this area of research include Baumgärtner et al. (1996), Beard and Lozada (1999), Burley and Foster (1994), Daly (1997a), Mayumi and Gowdy (1999), Pethig (2003) and Ruth (1999).

[16] Pethig (2003) surveys the Materials-Balance-Principle's origin and impact for environmental and resource economics.

2.4.3 Resource Extraction and Waste Generation

The insights described in Sections 2.4.1 and 2.4.2 have been applied, in particular, to the analysis of mineral resource extraction (e.g. Ruth 1995a, 1995b, 1995c), the generation of wastes and pollution (Kümmel 1989, Kümmel and Schüssler 1991), and the relation between the two (Faber 1985, Faber et al. 1995[1983]). At a very abstract level, high entropy (or: exergy lost) may be seen as the ultimate form of waste (Ayres et al. 1998, Ayres and Martinás 1995, Faber 1985, Kümmel 1989, Kümmel and Schüssler 1991).[17]

2.4.4 Representation of the Production Process

Every process of production is, at core, a transformation of energy and matter (Ayres and Kneese 1969). Hence, the laws of thermodynamics provide a suitable analytical framework for a rigorous deduction of insights into the physical aspects of production (Baumgärtner 2000). In particular, any representation of production in economic models should be in accordance with the laws of thermodynamics.[18] For this reason, the neoclassical production function, which is the standard way of representing the production process in economic models, has been critically discussed against the background of thermodynamics. It has become apparent that this concept is incompatible with the laws of thermodynamics for a number of reasons:

(i) Georgescu-Roegen (1971) claims that the neoclassical production function is incompatible with the laws of thermodynamics, basically, because it does not properly reflect the irreversible nature of transformations of energy and matter, and because it confounds flow and fund quantities (Daly 1997b, Kurz and Salvadori 2003).

(ii) One essential factor of production, which is very often omitted from the explicit representation, is energy (actually: exergy) (Ayres 1998, Kümmel 1989).[19] Its exact role for the production process, and its interplay with

---

[17]Waste materials deposited in the natural environment, however, might cause environmental problems not because of their high entropy, but precisely because their entropy is not yet maximal. In other words, it is the exergy still contained in waste materials, i.e. the potential to initiate chemical reactions and perform work, which makes these wastes potentially harmful to the natural environment (Ayres et al. 1998, Ayres and Martinás 1995, Perrings 1987). However, the view that the 'waste exergy' of by-products can be seen as a measure for potential harm done to natural ecosystems is limited. It does not take into account the (eco-)toxicity of some inert materials, nor does it take into account purely physical effects of inert materials, e.g. global warming due to the carbon dioxide emitted into the atmosphere.

[18]Krysiak and Krysiak (2003) discuss whether the neglect of the laws of thermodynamics in economics leads to any substantial problems.

[19]Econometric studies show that the production factor energy (exergy) explains a surprisingly large share of economic growth observed over the 20th century in the US, German or Japanese economies (Ayres et al. 2003, Kümmel et al. 1985, 2000).

other production factors, such as capital or material resources, is studied in engineering thermodynamics (e.g. Bejan 1996, 1997, Bejan et al. 1996; see also Section 2.4.5 below). This has lead to a new understanding of the role of energy for economic production processes, which goes beyond simply treating it as a factor of production (Buenstorf 2004).

(iii) The conservation laws for energy and matter imply that there are limits to substitution between energy-matter inputs, which are subject to the laws of thermodynamics, and other inputs such as labor or capital, which lie outside the domain of thermodynamics (Berry and Andresen 1982, Berry et al. 1978, Dasgupta and Heal 1979: Chapter 7).[20]

(iv) From the First and Second Laws of Thermodynamics it becomes obvious that '[g]iven the entropic nature of the economic process, waste is an output just as unavoidable as the input of natural resources' (Georgescu-Roegen 1975: 357). This holds not only for the economy at large, but for every individual process of production at the micro-level (Baumgärtner 2000: Chapter 5, 2002; Baumgärtner and de Swaan Arons 2003, Faber et al. 1998). As a result, there is no such thing as 'single production', i.e. the production of just one single output as modelled by the neoclassical production function. Rather, all production is joint production, i.e. there is necessarily more than one output (Baumgärtner et al. 2001, 2006, Faber et al. 1998).

All these apparent inconsistencies between the laws of thermodynamics and the standard assumptions about the neoclassical production function have led to more general descriptions of the production process, which blend the traditional theory of production with thermodynamic principles. Some of them use partial-analysis models (Anderson 1987, Ayres 1995, Ethridge 1973, Baumgärtner 2000: Chapter 4); others use general equilibrium models of the whole economy (Ayres and Kneese 1969, Perrings 1986, Krysiak and Krysiak 2003, Pethig 2006, Noll and Trijonis 1971).

2.4.5 Finite-Time/Finite-Size Thermodynamics: Exergy-Engineering

Recent research in the applied field of engineering thermodynamics has addressed the circumstance that chemical and physical processes in industry never

[20]Krysiak and Krysiak (2003) show that most abstract economic models, and general equilibrium theory in particular, are consistent with physical conservation laws. On the other hand, most applied models, also from environmental and resource economics, which use specific specifications of production functions, are not. As for the latter, Pethig (2006) shows that production functions with emissions treated as inputs can be reconstructed as a subsystem of a comprehensive production-cum-abatement technology that is in line with the materials balance principle.

happen in a completely reversible way between one equilibrium state and another equilibrium state. Rather, these processes are enforced by the operator of the process and they are constrained in space and time. This has led to an extension of ideal equilibrium thermodynamics, known as *finite-time/finite-size thermodynamics* (e.g. Andresen et al. 1984, Bejan 1996, 1997, Bejan et al. 1996).

From the point of view of finite-time/finite-size thermodynamics it becomes obvious that the minimum exergy requirement and minimum waste production in chemical or physical processes is considerably higher than that suggested by ideal equilibrium thermodynamics. The reason for the increased exergy requirement (which entails an increased amount of waste at the end of the process) lies in the fact that chemical and physical transformations are forced to happen over a finite time by the operator of the production plant, which necessarily causes some dissipation of energy.

The finite-time/finite-size consideration is a very relevant consideration for many production processes, in particular in the chemical industry. Finite-time/finite-size thermodynamics allows one to exactly identify, trace down and quantify exergetic inefficiencies at the individual steps of a production processes (Bejan 1996, 1997, Bejan et al. 1996, Brodyansky et al. 1994, Creyts 2000, Szargut et al. 1988), along the entire chain of a production process (Ayres et al. 1998, Cornelissen and Hirs 1999, Cornelissen et al. 2000), for whole industries (Dewulf et al. 2000, Hinderink et al. 1999, Ozdogan and Arikol 1981), and for entire national economies (Nakićenović et al. 1996, Schaeffer and Wirtshafter 1992, Wall 1987, 1990, Wall et al. 1994). Thus, it yields valuable insights into the origins of exergy losses and forms a tool for designing industrial production systems in an efficient and sustainable manner (Connelly and Koshland 2001, de Swaan Arons and van der Kooi 2001, de Swaan Arons et al. 2003).

Furthermore, it becomes apparent that energy/exergy and time are substitutes as factors of production in many production processes (Andresen et al. 1984; Berry and Andresen 1982, Spreng 1993). A production process may be speeded up at the expense of employing more energy/exergy, and the use of energy/exergy may be reduced by allowing the production process to just take longer. Prominent examples for such a trade-off-relationship are transport services or chemical reaction processes.

2.4.6 Thermodynamic and Economic Efficiency

Both thermodynamics and economics analyze systems in terms of their 'efficiency'. Both concepts may be applied to the very same system, e.g. a production plant or a whole national economy. Yet, the thermodynamic and the economic notions of efficiency differ fundamentally, as they refer to very different variables of the system. In fact, the two notions are completely independent (Berry et al. 1978, Dasgupta and Heal 1979: Chapter 7, Baumgärtner 2001).

As a result, thermodynamic efficiency is neither necessary nor sufficient for economic efficiency, even when economic efficiency includes concerns over energy, resources and environmental quality.

2.4.7 Sustainability: Limits to Economic Growth

From the very beginning, the recourse to thermodynamic arguments in ecological, environmental and resource economics was motivated by a long-term and global concern for the sustainable existence of humankind on 'Planet Earth' (Boulding 1966, Daly 1973, 1991[1977], Georgescu-Roegen 1971). The pre-analytic 'vision' (Schumpeter 1954: 42)[2] behind this concern was that of the human economy as an open subsystem of the larger, but finite, closed, and non-growing system of the biogeophysical environment.

In this view, thermodynamic analysis has helped to sketch the potential and limits of economic growth. It has turned out that limits to the growth of energy-matter throughput through the economy exist, which may ultimately set limits to economic growth. This claim is vindicated by the following arguments:[21]

(i) Conservation of mass implies that the marginal product as well as the average product of a material resource input may be bounded from above (Baumgärtner 2004a). This means that the usual Inada conditions (Inada 1963) do not hold for material resource inputs. This is important since the Inada conditions are usually held to be crucial for establishing steady state growth under scarce exhaustible resources (e.g. Dasgupta and Heal 1974, Solow 1974, Stiglitz 1974).

(ii) As described in Section 2.4.4 above, the conservation laws for energy and matter imply that there are limits to substitution between energy-matter inputs, which are subject to the laws of thermodynamics, and other inputs such as labor or capital, which lie outside the domain of thermodynamics (Berry and Andresen 1982, Berry et al. 1978, Dasgupta and Heal 1979: Chapter 7). This is important since substitutability among essential and scarce production factors (with an elasticity of substitution not smaller than one) is usually held to be crucial for establishing steady state growth (e.g. Dasgupta and Heal 1974, Solow 1974, Stiglitz 1974).

(iii) Some have posited that resource scarcity can be overcome by recycling. However, thermodynamic analysis clearly shows that there are limits to recycling as well (Ayres 1999b, Craig 2001).

(iv) Others still have posited that technical progress is an important driver of economic growth, and that technical progress will continue. However,

[21] Cleveland and Ruth (1997) present these arguments in more detail and review the relevant literature.

thermodynamic analysis clearly shows that there are limits to technical progress (Ruth 1995a, 1995b, 1995c).

2.5 Conclusion and Caveat: Thermodynamics and Sustainability

In conclusion, thermodynamic concepts, laws and models are relevant for ecological, environmental and resource economics in various ways and on different levels of abstraction.

(i) As all processes of change are essentially processes of energy and material transformation, the concepts and laws of thermodynamics apply to all of them. The framework of thermodynamics thus creates a unifying perspective on ecology, the physical environment, and the economy. This unifying framework, combined with economic and ecological analysis, allows asking questions which would not have been asked from the perspective of one scientific discipline alone.

(ii) On a more specific level thermodynamic concepts allow the incorporation of physical driving forces and constraints in models of economy-environment interactions, both microeconomic and macroeconomic. They are essential for understanding the extent to which resource and energy scarcity, nature's capacity to assimilate human wastes and pollutants, as well as the irreversibility of transformation processes, constrain economic action. Thermodynamic concepts thus allow economics to relate to its biogeophysical basis, and yield insights about that relation which are otherwise not available.

(iii) On an even more applied level, thermodynamic concepts provide tools for quantitative analysis of energetic and material transformations for engineers and managers. They may be used to design industrial production plants or individual components, such as to maximize their energetic efficiency and minimize their environmental impact.

With its rigorous but multifarious character as a method of analysis, its rich set of fruitful applications, and its obvious potential to establish relations between the natural world and purposeful human action, thermodynamics is therefore one of the cornerstones in the conceptual foundation of ecological, environmental and resource economics.

However, one important caveat seems to be in place. Thermodynamics is a purely descriptive science. That means, it only allows one to make statements of the kind 'If A, then B'. In particular, it is not a normative science. By itself, it neither includes nor allows value statements (Baumgärtner 2000: 65–66) or statements of the kind 'C is a good, and therefore desirable, state of the world, but D is not'.[22] In contrast, sustainability is essentially a normative issue (Faber et al. 1995, Faber et al. 1996: Chapter 5). Sustainability is about the question 'In what kind of world do we want to live today and in the future?', thus, inherently including a dimension of desirability. A purely descriptive science alone, like thermodynamics, cannot give an answer to that question.

Thermodynamics is necessary, however, to identify clearly the feasible options of development and their various properties, before a choice is then made about which option to choose based on some normative criteria. That choice requires a valuation or, more generally, a normative judgment of the different options at hand. It is therefore necessary not only to know the energetic and material basis of society's metabolism – both current and feasible alternatives – but also to link these thermodynamic aspects to the human perception and valuation of natural resources, commodity products and waste joint products, and the state of the natural environment.

The role of thermodynamics for conceiving sustainable modes of societal metabolism, therefore, is relative but essential. Thermodynamics is necessary to identify which options and scenarios of resource use, economic production, and waste generation are feasible and which are not. It thereby contributes to making informed choices about the future.

Appendix: Concepts and Laws of Thermodynamics

Thermodynamics is the branch of physics that deals with macroscopic transformations of energy and matter. Briefly summarized, the fundamental concepts and laws of phenomenological thermodynamics can be stated as follows.[23]

A2.1 Systems and Transformations

With respect to the potential exchange of energy and matter between the inside and the outside of the system under study, one distinguishes between the following types of thermodynamic system:

- *Isolated* systems exchange neither energy nor matter with their surrounding environment.

[22] This holds even for the notion of *thermodynamic efficiency*, which is a purely technical notion (see the discussion in Section 2.4.6 above).

[23] This appendix is taken from Baumgärtner (2002: Section 2.3). For a comprehensive introduction to (phenomenological) thermodynamics see Callen (1985), Kondepudi and Prigogine (1998) or Zemansky and Dittman (1997).

- *Closed* systems exchange energy, but not matter, with their surrounding environment.
- *Open* systems exchange both energy and matter with their surrounding environment.

A system is said to be in *thermodynamic equilibrium* when there is complete absence of driving forces for change in the system. Technically, the various potentials of the system are at their minimum, such that there are no spatial variations of any of the intensive variables within the system. *Intensive* variables are quantities which do not change when two separate but identical systems are coupled. In contrast, *extensive* variables are quantities whose value for the total system is simply the sum of the values of this quantity in both systems. For example, temperature and pressure are intensive variables while mass and volume are extensive ones. As long as there are spatial variations in, say, temperature within a system, it is not yet in thermodynamic equilibrium, because a potential for change exists. The equilibrium state is characterized by a uniform temperature throughout the system.

Consider an isolated system which undergoes a transformation over time between some initial equilibrium state and some final equilibrium state, either by interaction with its environment or by interaction between different constituents within the system. If the final state is such that no imposition or relaxation of constraints upon the isolated system can restore the initial state, then this process is called *irreversible*. Otherwise the process is called *reversible*. For example, at some initial time a gas is enclosed in the left part of an isolated box; the right part is separated from the left part by a wall and is empty. Now, the separating wall is removed. The molecules of the gas will then evenly distribute themselves over the entire volume of the box. The thermodynamic equilibrium of the final state is characterized by a uniform density of molecules throughout the entire volume. Reintroducing the wall into the isolated system separating the left part from the right half would not restore the initial state of the system. Nor would any other imposition or relaxation of constraints on the isolated system be able to restore the initial state. Therefore, the transformation given by the removal of the wall is an irreversible transformation of the isolated system.[24] Generally, a process of transformation can only be reversible if it does not involve any dissipation of energy, such as through e.g. friction, viscosity, inelasticity, electrical resistance or magnetic hysteresis.

---

[24]Note that this does not mean that the initial state of the system can never be restored. However, in order to restore the system's initial state, the initially isolated system has to be opened to the influx of energy. For instance, the initial state could be restored by removing the system's insulation and performing work on the system from the outside, e.g. by pressing all the molecules into the left part with a mobile wall that is initially at the right hand end of the system and from there on moves left.

A2.2 The Fundamental Laws of Thermodynamics

The *First Law of Thermodynamics* states that in an isolated system (which may or may not be in equilibrium) the total internal energy is conserved. This means that energy can be neither created nor destroyed. It can, however, appear in different forms, such as heat, chemical energy, electrical energy, potential energy, kinetic energy, work, etc. For example, when burning a piece of wood or coal the chemical energy stored in the fuel is converted into heat. In an isolated system the total internal energy, i.e. the sum of energies in their particular forms, does not change over time. In any process of transformation only the forms in which energy appears change, while its total amount is conserved.

Similarly, in an isolated system the total mass is conserved (*Law of Conservation of Mass*). Obviously, if matter cannot enter or leave an isolated system, the number of atoms of any chemical element within the system must remain constant. In an open system which may exchange matter with its surrounding, a simple *Materials Balance Principle* holds: the mass content of a system at some time is given by its initial mass content plus inflows of mass minus outflows of mass up to that point in time. The law of mass conservation, while often regarded as an independent conservation law besides the law of energy conservation, is actually an implication of the First Law of Thermodynamics. According to Einstein's famous relation $E = mc^2$ mass is a form of energy, but mass can only be transformed into non-material energy, and vice versa, in nuclear reactions. Therefore, neglecting nuclear reactions, it follows from the First Law of Thermodynamics that mass and non-material energy are conserved separately.

In any process transforming energy or matter, a certain amount of energy is irrevocably transformed into heat. The variable *entropy* has been defined by Rudolph Clausius (1854, 1865) such as to capture this irrevocable transformation of energy: if a certain amount of heat dQ is reversibly transferred to or from a system at temperature T, then $dS = dQ/T$ defines the entropy S. Clausius showed that S is a state variable of the system, i.e. it remains constant in any reversible cyclic process, and increases otherwise. The *Second Law of Thermodynamics*, the so-called Entropy Law, states the unidirectional character of transformations of energy and matter: With any transformation between an initial equilibrium state and a final equilibrium state of an isolated system, the entropy of this system increases over time or remains constant. It strictly increases in irreversible transformations, and it remains constant in reversible transformations, but it cannot decrease.

Entropy, in this view, can be interpreted as an indicator for the system's capacity to perform useful work. The higher the value of entropy, the higher the amount of energy already irreversibly transformed into heat, the lower the amount of free energy of the system and the lower the system's capacity to perform work. Expressed the other way round, the lower the value of entropy, the higher the amount of free energy in the system and the higher the system's capacity to perform work. Hence, the statement of the Second Law of Thermodynamics amounts to saying that, for any process of transformation, the proportion of energy in the form of heat to total energy irreversibly increases or remains constant, but certainly never decreases. In other words, with any transformation of energy or matter, an isolated system loses part of its ability to perform useful mechanical work and some of its available free energy is irreversibly transformed into heat. For that reason, the Second Law is said to express an irreversible degradation of energy in isolated systems over time. Through this, the economic relevance of the Second Law becomes obvious.

While the notion of entropy introduced to phenomenological thermodynamics by Clausius is based on heat, Ludwig Boltzmann (1877) introduced a formally equivalent notion of entropy that is based on statistical mechanics and likelihood. His notion reveals a different interpretation of entropy and helps to show why it irreversibly increases over time. Statistical mechanics views gases as assemblies of molecules, described by distribution functions depending on position and velocity. This view allows the establishment of connections between the thermodynamic variables, i.e. the macroscopic properties such as temperature or pressure, and the microscopic behaviour of the individual molecules of the system, which is described by statistical means.[25] The crucial step is to distinguish between microstates and macrostates of a system. The *microstate* is an exact specification of the positions and velocities of all individual particles; the *macrostate* is a specification of the thermodynamic variables of the whole system.

Boltzmann assumed that all microstates have equal *a priori* probability, provided that there is no physical condition which would favour one configuration over the other. He posited that every macrostate would always pass to one of higher probability, where the probability of a macrostate is determined by the number of different microstates realising this macrostate. The macroscopic thermal equilibrium state is then the most probable state, in the sense that it is the macrostate which can be realized by the largest number of different microstates. Boltzmann defined the quantity Ω, counting the number of possible microstates realising one macrostate, and related this to the thermodynamic entropy S of that macrostate. He used $S = k \log \Omega$, with k as a factor of proportionality called Boltzmann's constant. Entropy can thus be taken as a measure of likelihood: highly probable macrostates, that is macrostates which can be realized by a large number of microstates, also have high entropy. At the same time, entropy may be interpreted as a measure of how orderly or mixed-up a system is. High entropy, according to the Boltzmann interpretation, characterizes a system in which the individual constituents are arranged

[25]Balian (1991), Huang (1987) and Landau and Lifshitz (1980) give an introduction to statistical mechanics.

in a spatially even and homogeneous way ('mixed-up systems'), whereas low entropy characterizes a system in which the individual constituents are arranged in an uneven and heterogeneous way ('orderly systems'). The irreversibility stated by the Second Law in its phenomenological formulation (in any isolated system entropy always increases or remains constant) now appears as the statement that any isolated macroscopic system always evolves from a less probable (more orderly) to a more probable (more mixed-up) state, where Ω and S are larger.

Whereas the Second Law in its Clausius or Boltzmann formulation makes a statement about isolated systems in thermodynamic equilibrium only, the study of closed and open systems far from equilibrium has shown (Prigogine 1962, 1967) that entropy is also a meaningful and useful variable in closed and open systems. Any open system is a subsystem of a larger and isolated system. According to the conventional formulation of the Second Law, the entropy of the larger and isolated system has to increase over time, but the entropy of any open subsystem can, of course, decrease. Viewing open systems as subsystems of larger and isolated systems reveals, however, that an entropy decrease in an open subsystem necessarily has to be accompanied by an entropy increase in the system's environment, that is the rest of the larger, isolated system, such that the entropy of the total system increases.

A generalization of the Second Law is possible so that it does not only refer to isolated systems. Irrespective of the type of thermodynamic system under study, and irrespective of whether the system is in thermodynamic equilibrium or not, it is true that entropy cannot be annihilated; it can only be created (Falk and Ruppel 1976: 353). This more general, system-independent formulation of the Second Law implies the usual formulations for isolated systems. The relevance of the system-independent formulation of the Second Law lies in the fact that most real systems of interest are not isolated but closed or open. Hence, the latter formulation is the form in which the Second Law is apparent in everyday life.

A2.3 Quantification and Application

The entropy concept is essential for understanding how resource and energy scarcity, as well as the irreversibility of transformation processes, constrain economic action (Baumgärtner 2003, Georgescu-Roegen 1971). However, it is a very abstract concept and it is notoriously difficult to apply in specific contexts. One of the complications is due to the fact that a system's capacity to perform work depends not only on the state of the system, but also on the state of the system's environment. Therefore, for applications of the fundamental thermodynamic insights in the areas of mechanical and chemical engineering, as well as in economics, it is useful to relate the system's ability to perform work to a certain standardized reference state of its environment. *Exergy* is defined to be the maximum amount of work obtainable from a system as it approaches thermodynamic equilibrium with its environment in a reversible way (Szargut et al. 1988: 7). Exergy is also commonly called *available energy*, or *available work*, and corresponds to the 'useful' part of energy, thus combining the insights from both the First and Second Laws of Thermodynamics. Hence, exergy is what most people mean when they use the term 'energy' carelessly, e.g. when saying that 'energy is used' to carry out a certain process.

The relationship between the concepts of entropy and exergy is simple, as $B_{\text{lost}} = T_0 S_{\text{gen}}$ (*Law of Gouy and Stodola*), where B_{lost} denotes the potential work or exergy lost by the system in a transformation process, T_0 denotes the temperature of the system's environment, and S_{gen} denotes the entropy generated in the transformation. This means, as the system's entropy increases as a consequence of irreversible transformations according to the Second Law, the system loses exergy or some of its potential to perform work. Exergy, unlike energy, is thus not a conserved quantity. While the entropy concept stresses that with every transformation of the system something useless is created, the exergy concept stresses that something useful is diminished. These developments are two aspects of the same irreversible character of transformations of energy and matter.

As the system might consist simply of a bulk of matter, exergy is also a measure for the potential work embodied in a material, whether it is a fuel, food or other substance (Ayres 1998, Ayres et al. 1998). The exergy content of different materials can be calculated for standard values specifying the natural environment, by considering how that material eventually reaches thermodynamic equilibrium with its environment with respect to temperature, pressure, chemical potential and all other intensive variables.[26] While taking a particular state of the system's environment as a reference point for the definition and calculation of exergy may be considered a loss of generality as compared to the entropy concept, this referencing seems permissible since all processes of transformation – be it in nature or in the economy – are such that:

(i) all the materials involved eventually do reach thermodynamic equilibrium with the natural environment;

(ii) the environment is so large that its equilibrium will not be affected by the particular transformation processes under study.

While both the entropy and the exergy concept yield the same qualitative insights into the fundamentally irreversible character of transformations of energy and matter, the exergy concept is more tangible, as it is directly related to the

[26] Exergy values for many materials are typically calculated for an environmental temperature of 298.15 K and pressure of 101.325 kPa and can be found in tables, such as e.g. in Szargut et al. (1988: Appendix).

very compelling idea of 'available work' and can be more easily quantified than entropy.

3. The Inada Conditions for Material Resource Inputs Reconsidered*

3.1 Introduction

It is characteristic for many of the pioneering theoretical contributions to the analysis of economic growth under scarcity of exhaustible natural resources (Dasgupta and Heal 1974, Hoel 1978, Mäler 1974, Schulze 1974, Smith 1977, Solow 1974, Stiglitz 1974, Weinstein and Zeckhauser 1974) as well as large parts of the vast strand of literature which they initiated that they take the well established neoclassical growth theory as their starting point and extend it in the most simple way to also include natural resources, namely by adding one additional variable representing material resource input into a neoclassical aggregate production function. This production function is usually assumed to display the standard properties concerning the resource factor (substitutability between factors of production, positive decreasing marginal product approaching zero and infinity in the two limits of infinite and vanishing resource input).[1]

However, this procedure does not appropriately account for the fact that the extraction of material resource inputs, their transformation within the production process, and their emission or disposal after use are, at root, transformations of energy and matter. As such they are subject to the laws of thermodynamics, which is the branch of physics dealing with transformations of energy and matter. Thermodynamic relations, thus, may impose additional constraints on economic action (Daly 1997b, Solow 1997).

This chapter formally explores one particular implication that the thermodynamic law of conservation of mass, the so-called Materials-Balance-Principle, has for modeling production. It is shown that the marginal product as well as the average product of a material resource input are bounded from above. This

*Previously published in *Environmental and Resource Economics*, **29**(3), 307–322 (2004).

[1]Historically, the concept of a production function has been introduced by Wicksell (1893) and Wicksteed (1992[1894]) to analyze the *distribution* of income among the factor owners, and not its *physical production* (Schumpeter 1954: 1028, Sandelin 1976). But later on, the concept has come to dominate economists' thinking about the physically feasible production possibilities, as in the discussion of economic growth under scarcity of natural resources.

means that the usual Inada conditions (Inada 1963), when applied to material resource inputs, are inconsistent with a basic law of nature. This is important since the Inada conditions are usually held to be crucial for establishing steady state growth under scarce exhaustible resources. While the advocates of a thermodynamic-limits-to-economic-growth perspective (e.g. Boulding 1966, Daly 1991[1977], Georgescu-Roegen 1971) usually stress the universal and inescapable nature of limits imposed by laws of nature, pro-economic-growth advocates usually claim that there is plenty of scope for getting around particular thermodynamic limits by substitution, technical progress and 'dematerialization' (e.g. Beckerman 1999, Smulders 1999, Stiglitz 1997). The latter therefore often conclude that, on the whole, thermodynamic constraints are simply irrelevant for economics. This chapter takes a more differentiated stand, by analyzing in detail

(i) what exactly are the implications of thermodynamics for modeling production at the level of a single production process, and

(ii) how these constraints carry over to the level of aggregate production, considering that there is scope for substitution in an economy between different resources and different production technologies.

This chapter continues and merges two strands in the literature on the production function. In a first strand, the neoclassical production function has been critically discussed against the background of thermodynamics. Georgescu-Roegen (1971) claims that the neoclassical production function is incompatible with the laws of thermodynamics, basically because it does not properly reflect the irreversible nature of transformations of energy and matter, and because it confounds flow and fund quantities (Daly 1997a, Kurz and Salvadori 2003). Berry et al. (1978) and Dasgupta and Heal (1979: Chapter 7) demonstrate that the conservation laws for energy and matter imply that substitutability between energy-matter inputs, which are subject to the laws of thermodynamics, and other inputs such as labor or capital, which lie outside the domain of thermodynamics, is restricted. All these apparent inconsistencies between the laws of thermodynamics and the standard assumptions about the neoclassical production function have led to more general descriptions of the production process, which blend the traditional theory of production with the thermodynamic principle of conservation of mass (Anderson 1987, Baumgärtner 2000: Chapter 4, Pethig 2003: Section 3.3).

Another strand in the literature, more narrowly concerned with production theory (Shephard 1970), has focused specifically on the Inada conditions. It has been demonstrated that the Inada conditions follow from other basic properties of the neoclassical production function (Dyckhoff 1983), and that they impose strong restrictions on the asymptotic behavior of the elasticity of substitution between capital and labor (Barelli and de Abreu Pessôa 2003). Furthermore, the Inada conditions have been shown to be incompatible with another basic principle within economics, the Law of Diminishing Returns (Färe and Primont 2002).

The chapter is organized as follows. In Section 3.2, the Inada conditions are briefly reviewed in the context of neoclassical growth theory with and without natural resources. Section 3.3 provides a thermodynamic analysis of the production process at the micro level, i.e. for a micro production function for a single commodity. Section 3.4 explores the implications for the Inada conditions at the macro-level, i.e. for an aggregate production function for an all purpose commodity. Section 3.5 concludes.

3.2 The Inada Conditions on Resource Inputs

With just two inputs, capital K and labor L, the aggregate neoclassical production function for output Y takes the form $Y = F(K, L)$. It is usually assumed to exhibit constant returns to scale and positive and diminishing marginal products with respect to each input for all $K, L > 0$ (Solow 1956, Swan 1956):

$$\frac{\partial F}{\partial K} > 0, \quad \frac{\partial^2 F}{\partial K^2} < 0, \quad \frac{\partial F}{\partial L} > 0, \quad \frac{\partial^2 F}{\partial L^2} < 0. \tag{3.1}$$

Furthermore, following Inada (1963) the marginal product of an input is assumed to approach infinity as this input goes to zero and to approach zero as the input goes to infinity:

$$\lim_{K \to 0} \frac{\partial F}{\partial K} = \lim_{L \to 0} \frac{\partial F}{\partial L} = +\infty, \qquad \lim_{K \to +\infty} \frac{\partial F}{\partial K} = \lim_{L \to +\infty} \frac{\partial F}{\partial L} = 0. \tag{3.2}$$

In growth models these so-called Inada conditions are crucial for the existence of interior steady state growth paths: they are sufficient (yet not necessary) for the existence of an interior solution in which the economy grows at a constant and strictly positive rate. Assumptions (3.1) and (3.2) imply that each input is essential for production, that is, $F(0, L) = F(K, 0) = 0$, and that output goes to infinity as either input goes to infinity.

When extending the framework of neoclassical growth theory to also include natural resources this is usually done by including one additional variable, R, into the production function, representing material resource input: $Y = F(K, L, R)$ (Dasgupta and Heal 1974: 9, Solow 1974: 34, Stiglitz 1974: 124). The same standard assumptions are then made about this resource dependent production function F as made before about the capital-labor-only-production function. For instance, F is assumed to be increasing, strictly concave, twice differentiable, and linear homogeneous (Dasgupta and Heal 1974: 9, Solow 1974: 34, Stiglitz 1974: 124). Furthermore, some more or less direct analogue to the Inada conditions is assumed in order to guarantee existence of non-trivial (interior) solutions. For example, Solow (1974: 34) assumes that resources are essential for production, i.e. $F(K, L, 0) = 0$, and, at the same time the average product of R has no upper bound, i.e. there is no $\alpha < +\infty$ with $F/R < \alpha$. While this is a particular form of the Inada condition, since it necessarily follows from $\lim_{R \to 0} \partial F/\partial R = +\infty$, Dasgupta and Heal (1974: 11) directly assume that $\lim_{R \to 0} \partial F/\partial R = +\infty$.

Based on one or the other form of Inada conditions, the result is that even with a limited reservoir of an exhaustible natural resource and with that resource being essential for production it is possible to maintain a positive and constant level of consumption forever (Solow 1974). If there is technical progress there might even be exponentially growing consumption (Stiglitz 1974). And while the remaining stock of the resource will approach zero along the optimal path, the resource will never completely be exhausted (Dasgupta and Heal 1974).

So, in a sense, these analyses seem to have produced a rather optimistic answer to the 'Limits to growth'-concern (Meadows et al. 1972). However, the Inada conditions as applied (in whatever form) to material resource inputs may be inconsistent with the thermodynamic law of conservation of mass. This is demonstrated in the following.

3.3 Thermodynamic Limits to Resource Productivity at the Micro Level

The First Law of Thermodynamics implies that matter can neither be created nor annihilated, i.e. in a closed system it is conserved.[2] This law establishes what is known as 'Materials-Balance-Principle' in environmental and resource economics (Ayres 1999a, Pethig 2003).

In order to infer this Law's implications for the production process, consider the following simple model of production at the micro level, i.e. production of a particular good by a particular elementary technology. For the moment assume that there is only one single natural resource.[3] Production of output Y depends – besides capital K and labor L – on the resource material R:

$$Y = F(K, L, R). \tag{3.3}$$

[2] A *closed* thermodynamic system is one that does not exchange matter with its surrounding. It may, however, exchange energy with its surrounding. For an elementary introduction into thermodynamics, see e.g. Kondepudi and Prigogine (1998).

[3] The generalization to many different natural resources will be done in Section 3.4 below.

As a by-product the production process yields the non-negative amount W of waste. All three, R, Y and W, are measured in physical (mass) units. Let ρ with $0 < \rho \leq 1$ denote the (mass) fraction of resource material contained in the output, and μ with $0 \leq \mu \leq 1$ the (mass) fraction of resource material contained in the waste.[4] While ρ is, in general, a function of K and L, i.e. the resource content of the final product may be decreased by using more capital and labor ('dematerialization'), there obviously are physical limits to dematerialization. For instance, in order to produce one kilogram of iron screws one needs to employ at least one kilogram of pure iron. This means that ρ is physically bounded from below, in particular $\rho > 0$. Therefore, one may take ρ as a constant denoting the lower bound to dematerialization, i.e. $\rho = const.$ with $0 < \rho \leq 1$ denotes the *minimum* (mass) fraction of resource material contained in the output.

Applying the materials-balance-principle to the production process results in the following formal balance equation:

$$R = \rho F(K, L, R) + \mu W, \tag{3.4}$$

which states that the resource material which enters the process also eventually has to come out of the process, be it in the desired product or in the waste. Rearranging Equation (3.4) into

$$\frac{F(K, L, R)}{R} = \frac{1}{\rho}\left(1 - \frac{\mu W}{R}\right)$$

and noting that $\mu W/R \geq 0$ yields the following upper bound for the average resource product F/R:

$$\frac{F(K, L, R)}{R} \leq \frac{1}{\rho}. \tag{3.5}$$

This establishes the following result.

Proposition 3.1
The average product of resource input, $F(K, L, R)/R$, is bounded from above by the inverse of the resource fraction in the good produced, $1/\rho$.

Proposition 3.1 has the following implication for the shape of the production function. Equation (3.5) can be rearranged into

$$F(K, L, R) < \frac{1}{\rho} R,$$

[4] That ρ and μ are allowed to be less than 1 is due to materials other than the natural resource R considered here. These other materials might enter the production process and be part of the product as well as of the waste. Yet, they are not explicitly represented here.

which states that for fixed values of K and L the graph of F plotted as a function of R always stays below a line of slope $1/\rho$ starting at the origin (Figure 3.1). As ρ becomes smaller, the upper limit on the average resource product will grow. However, the upper limit will always remain finite, since ρ is strictly positive.

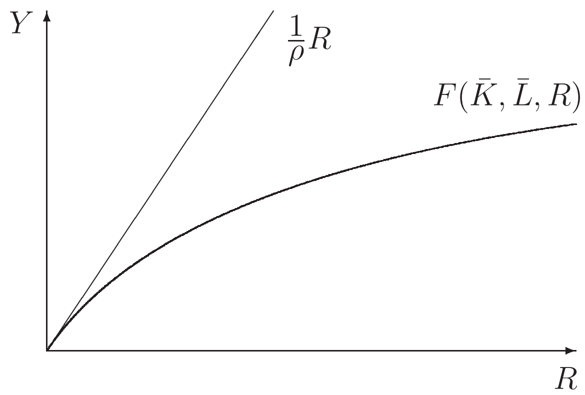

Figure 3.1 The materials-balance-principle implies that the graph of $F(\bar{K}, \bar{L}, R)$ is bounded from above by a line of slope $1/\rho$ starting at the origin.

With the *average* resource product F/R bounded from above by the inverse of the resource fraction in the good produced, $1/\rho$, a similar argument applies to the *marginal* resource product. Taking the total differential of the material balance Equation (3.4) and considering only variations in R, i.e. $dK = dL = 0$, yields

$$dR = \rho \, \frac{\partial F(K, L, R)}{\partial R} \, dR + \mu \, \frac{\partial W(K, L, R)}{\partial R} \, dR \; .$$

This holds for all $dR \geq 0$ and, thus, implies

$$1 = \rho \, \frac{\partial F(K, L, R)}{\partial R} + \mu \, \frac{\partial W(K, L, R)}{\partial R} \; . \tag{3.6}$$

Equation (3.6) simply is the materials-balance-equation for one additional unit of resource input employed in the production process. It leaves the process either as part of the desired product or as waste. The amount of the latter, $\partial W/\partial R$, obviously, cannot be negative. Therefore, rearranging Equation (3.6) into

$$\frac{\partial F(K, L, R)}{\partial R} = \frac{1}{\rho} \left(1 - \mu \, \frac{\partial W(K, L, R)}{\partial R} \right)$$

and noting that $\mu \partial W/\partial R \geq 0$ yields the following upper bound for the marginal resource product $\partial F/\partial R$:

$$\frac{\partial F(K,L,R)}{\partial R} \leq \frac{1}{\rho}. \tag{3.7}$$

This establishes the following result.

Proposition 3.2
The marginal product of resource input, $\partial F/\partial R$, is bounded from above by the inverse of the resource fraction in the good produced, $1/\rho$.

It is immediately obvious that if the marginal resource product F_R is bounded from above by $1/\rho$, then the marginal resource product as resource input approaches zero is also bounded from above be the same value:

$$\lim_{R \to 0} \frac{\partial F(K,L,R)}{\partial R} \leq \frac{1}{\rho}\ . \tag{3.8}$$

This is the content of the following corollary to Proposition 3.2.

Corollary
The marginal product of resource input as resource input approaches zero is bounded from above by the inverse of the resource fraction in the good produced, $1/\rho$.

By this corollary it becomes apparent that the Inada conditions (in whatever form), when applied to a micro level production function with material resource inputs, are inconsistent with the Materials-Balance-Principle.

The intuition behind the simple formal exercise carried out in this section is that matter cannot be created and, consequently, the produced output cannot contain more of some material than has been supplied as input. If, for instance, one needs 100 gram of some resource material in order to produce 1 kilogram of a good ($\rho = 1/10$), then, out of 1 kilogram of the resource one can produce at maximum (i.e. with no waste) 10 kilogram of output. This means, the average as well as the marginal resource product is bounded from above by 10 ($= 1/\rho$).

3.4 Thermodynamic Limits to Resource Productivity at the Macro Level

The simple model of production specified by Equation (3.3) was confined to the description of one particular *micro level* production process and one particular natural resource. In order to analyse how the thermodynamic law of conservation of mass may restrict *aggregate* production, we should think of production in a more general way:

- There are many different natural resources, such that one can substitute from one resource to another, in order to avoid thermodynamic constraints on micro level production, such as Conditions (3.5) or (3.7), to become binding.
- Production of an aggregate output, such as an all purpose commodity or GDP, is a multi-level process. On a first level, a number of different intermediate goods are produced from elementary resources (*micro level production*). On a second level, the final output is produced from the intermediate goods (*macro level production*).[5]
- In the production of final output there is scope for substitution between the input of different intermediate goods.

In such a setting, there is plenty of scope for substitution both between different elementary resource materials and between production processes. The question then is: To what extent do thermodynamic constraints on micro level production processes, such as Conditions (3.5) or (3.7), carry over to the macro level? And how do the laws of thermodynamics restrict aggregate production in such a general setting?

To answer these questions, consider the following model of production of an all purpose commodity from intermediate goods, which are themselves produced from a variety of elementary natural resources. There are n different elementary natural resources, numbered by $i = 1, \ldots, n$. Assume that this is a complete and exhaustive list of material resources actually or potentially used in production. For example, one may think of this list of natural resources as the complete list of known chemical elements, in which case $n = 112$.[6] These are used as inputs in the production of m different intermediate goods, each of which is produced by a single process of production, numbered by $j = 1, \ldots, m$. Production of theses intermediates is described by production functions

$$Y_j = F^j(K_j, L_j, R_{1j}, \ldots, R_{nj}) \qquad \text{for all } j = 1, \ldots, m, \tag{3.9}$$

where K_j and L_j denote input of capital and labor into production of intermediate good j. Similarly, R_{ij} (with $i = 1, \ldots, n$ and $j = 1, \ldots, m$) denotes input of resource material i into production of intermediate good j. Then,

$$R_i = \sum_{j=1}^{m} R_{ij} \tag{3.10}$$

[5]In general, aggregate production may be over more than two levels. But the essential insights can already be grasped from considering a two-level-system.

[6]As of 2003, there are 112 known chemical elements, 83 of which are naturally occurring. Examples include hydrogen, carbon, oxygen, iron, copper, aluminum, gold and uranium. Elements 113 through 118 are known to exist, but are not yet discovered (IUPAC 2003).

is the total amount of resource material i utilized in production. Each production process F^j also yields a certain amount of waste, W_j.

Let ρ_{ij} with $0 \leq \rho_{ij} \leq 1$ denote the (mass) fraction of resource material i contained in the intermediate good j, and μ_{ij} with $0 \leq \mu_{ij} \leq 1$ the (mass) fraction of resource material i contained in the waste from producing intermediate good j. Note that ρ_{ij} will be zero if the intermediate good j does not contain any material of type i. This may include cases in which some resource material has been used in, or is even essential for, the production of the intermediate, say as a catalyst, yet the material is not contained in the good produced. Nonetheless, every intermediate good j – as long as it is a material good and not an immaterial service – contains *some* amount of *some* of the materials, while not containing anything of other materials. In order to make this distinction explicit, let

$$I_j = \{i \,|\, \rho_{ij} > 0\} \subseteq \{1, \ldots, n\} \tag{3.11}$$

be the set of all resources which make up – as far as material content goes – the intermediate good j. The complement set $\bar{I}^j = \{1, \ldots, n\} \setminus I^j$ then denotes the set of all resources which are not materially contained in the intermediate good j.

The final good, an all purpose commodity, is produced from capital K, labor L and the intermediate goods Y_j (with $j = 1, \ldots, m$):

$$Y = F(K, L, Y_1, \ldots, Y_m), \tag{3.12}$$

where Y_j (with $j = 1, \ldots, m$) denotes input of intermediate good j as produced on the first level of production (Equation 3.9). On this level, elementary resources do not enter directly into production, but only indirectly insofar as they are embedded in the intermediates.[7] The final good production function (3.12) can be interpreted as an aggregate production function of the economy, specifying how the final good is produced from elementary resources, when the Y_j's are replaced by the respective micro level production functions (Equation 3.9).

The production of the final good also yields a certain amount of waste, W. Let ρ_i with $0 \leq \rho_i \leq 1$ denote the (mass) fraction of resource material i contained in the final output, and μ_i with $0 \leq \mu_i \leq 1$ the (mass) fraction of the resource material i contained in the waste. Note that ρ_i will be zero if the final good does not contain any material of type i. Nevertheless, the final good – as long as it is a material good and not an immaterial service – contains *some* amount of *some* of the materials, while not containing anything of other materials. For example, a passenger car may contain aluminum, carbon and

[7]This assumption, which is also quite plausible, only serves to simplify the notation and does not restrict the validity of results. It could easily be relaxed.

thallium, but no gold or plutonium. In order to make this distinction explicit, let

$$I = \{i \,|\, \rho_i > 0\} \subseteq \{1, \ldots, n\} \tag{3.13}$$

be the set of all resources which make up – as far as material content goes – the final good. The complement set $\bar{I} = \{1, \ldots, n\} \setminus I$ then denotes the set of all resources which are not materially contained in the final good. Assume that the final good is material, that is, it contains at least one type of material.

Assumption 3.1
I is non-empty.

With this setting and notation, Propositions 3.1 and 3.2 derived in Section 3.3 above can obviously be translated and generalized into the following statement:

Lemma 3.1
The thermodynamic law of conservation of mass implies that the micro level production functions $F^j(K_j, L_j, R_{1j}, \ldots, R_{nj})$ for all $j = 1, \ldots, m$ have the following properties:

$$\begin{aligned} \frac{F^j(K_j, L_j, R_{1j}, \ldots, R_{nj})}{R_{ij}} &\leq \frac{1}{\rho_{ij}} && \textit{and} \\ \frac{\partial F^j(K_j, L_j, R_{1j}, \ldots, R_{nj})}{\partial R_{ij}} &\leq \frac{1}{\rho_{ij}} && \textit{for all } i \in I_j . \end{aligned}$$

In words, the average and marginal resource product of resource material i in producing the intermediate good j is bounded from above by $1/\rho_{ij}$ in all cases where material i is contained in intermediate good j. If, in contrast, material i is not contained in intermediate good j, the average and marginal resource product of resource material i do not need to be bounded from above.[8]

Considering the overall two-level production system, the formal balance conditions for resource material i (for all $i = 1, \ldots, n$) then read as follows:

$$R_i = \sum_{j=1}^{m} R_{ij} \; , \tag{3.14}$$

$$R_{ij} = \rho_{ij} F^j(K_j, L_j, R_{1j}, \ldots, R_{nj}) + \mu_{ij} W_j \quad \text{for all } j = 1, \ldots, m \; , \tag{3.15}$$

$$\sum_{j=1}^{m} \rho_{ij} F^j(K_j, L_j, R_{1j}, \ldots, R_{nj}) = \rho_i F(K, L, Y_1, \ldots, Y_m) + \mu_i W \; . \tag{3.16}$$

[8] Of course, in the latter case they may still be bounded from above for reasons other than thermodynamic necessity.

Equation (3.14) states that the total amount of resource material i employed in production, R_i, may be used in any of the m production processes for intermediate goods. Equation (3.15) expresses conservation of mass of resource material i on the first level of production in all of the m intermediate good production processes: the total amount of material utilized in one of these processes, R_{ij}, leaves the process either as part of the intermediate good j or as part of the waste generated by that process. Equation (3.16) expresses conservation of mass on the second level of production: resource material i enters production of the final product indirectly, namely embedded in the m intermediate goods, each of which has a material content of that material of $\rho_{ij} Y_j$. It leaves the production process either as part of the final good or as part of the waste generated by that process. Summing balance Equation (3.15) over all micro level processes j, adding balance Equation (3.16) for the macro level, and using (3.14) yields an overall balance condition for material i:

$$R_i = \rho_i F(K, L, Y_1, \ldots, Y_m) + \mu_i W + \sum_{j=1}^{m} \mu_{ij} W_j \ .$$

This condition states that the material utilized in production, R_i, leaves the two-level production system either as part of the final good, or as part of the waste generated by the final good production process, or as part of the waste generated in any of the m intermediate good production processes.

Rearranging Equation (3.17) into

$$\frac{F(K, L, Y_1, \ldots, Y_m)}{R_i} = \frac{1}{\rho_i} \left[1 - \frac{\mu_i W}{R_i} - \frac{\sum\limits_{j=1}^{m} \mu_{ij} W_j}{R_i} \right] \tag{3.17}$$

and noting that $\mu_i W / R_i \geq 0$ as well as $\sum_{j=1}^{m} \mu_{ij} W_j / R_j \geq 0$, the following inequality holds:

$$\frac{F(K, L, Y_1, \ldots, Y_m)}{R_i} \leq \frac{1}{\rho_i}. \tag{3.18}$$

For all materials $i \in I$ which make up the final good, ρ_i is strictly positive so that $1/\rho_i < +\infty$ is a a finite upper bound for the average resource product of material i in aggregate production, F/R_i. From Assumption 3.1 it follows that there is at least one such material. For all other materials with $i \notin I$, ρ_i is zero so that $1/\rho_i$ is not a finite upper bound. This establishes the following result.

Proposition 3.3

(i) For all materials $i \in I$, which make up the final good, the average product of resource material i in aggregate production, F/R_i, is bounded from above by the inverse of this material's mass fraction in the final good, $1/\rho_i$.

(ii) *There exists at least one such material for which the average product is bounded from above.*

(iii) *For all materials $i \notin I$, which are not contained in the final good, the average product of resource material i in aggregate production, F/R_i, does not need to be bounded from above.*

In order to derive an analogue statement about the marginal resource products, take the total differential of the material balance Equation (3.17) for material i, with the Y_j in production function F replaced my the intermediate good production functions F^j (Equation 3.9), and consider only variations in resource material i (i.e. $dK = dL = 0$, $dK_j = dL_j = 0$ for all j and $dR_{i'j} = 0$ for all $i' \neq i$):

$$dR_i = \rho_i \sum_{j=1}^{m} \frac{\partial F}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} dR_{ij} + \mu_i \sum_{j=1}^{m} \frac{\partial W}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} dR_{ij} + \sum_{j=1}^{m} \mu_{ij} \frac{\partial W_j}{\partial R_{ij}} dR_{ij} \ . \quad (3.19)$$

From balance Equation (3.14) it follows that

$$dR_i = \sum_{j=1}^{m} dR_{ij} \ . \quad (3.20)$$

Replacing dR_i in Equation (3.19) by expression (3.20) and rearranging terms yields

$$\sum_{j=1}^{m} \left[1 - \rho_i \frac{\partial F}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} - \mu_i \frac{\partial W}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} - \mu_{ij} \frac{\partial W_j}{\partial R_{ij}} \right] dR_{ij} = 0 \ . \quad (3.21)$$

This holds for all $dR_{ij} \geq 0$ and, thus, implies that the term in brackets equals zero. This can be rearranged into

$$\frac{\partial F}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} = \frac{1}{\rho_i} \left[1 - \mu_i \frac{\partial W}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} - \mu_{ij} \frac{\partial W_j}{\partial R_{ij}} \right] \ . \quad (3.22)$$

Noting that the second and third term in brackets are non-negative yields the following inequality, which holds for all i and j:

$$\frac{\partial F}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} \leq \frac{1}{\rho_i} \ . \quad (3.23)$$

On the other hand, taking the total differential of the defining equation for production function F (Equation 3.12), with the Y_j in production function F replaced my the intermediate good production functions F^j (Equation 3.9), and considering only variations in resource material i (i.e. $dK = dL = 0$, $dK_j = dL_j = 0$ for all j and $dR_{i'j} = 0$ for all $i' \neq i$), yields:

$$dF = \sum_{j=1}^{m} \frac{\partial F}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} \, dR_{ij} \; . \tag{3.24}$$

From (3.24) one obtains, using Inequality (3.23) and Equation (3.20)

$$dF = \sum_{j=1}^{m} \frac{\partial F}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} \, dR_{ij} \leq \sum_{j=1}^{m} \frac{1}{\rho_i} \, dR_{ij} = \frac{1}{\rho_i} \sum_{j=1}^{m} dR_{ij} = \frac{1}{\rho_i} \, dR_i \; , \tag{3.25}$$

so that we have the following inequality:

$$dF \leq \frac{1}{\rho_i} \, dR_i \; . \tag{3.26}$$

Interpreting this inequality for differentials as an algebraic expression and rearranging finally yields:

$$\frac{dF}{dR_i} \leq \frac{1}{\rho_i}. \tag{3.27}$$

Since the production function F (Equation 3.12) does neither directly nor indirectly depend on R_i, the expression dF/dR_i should not be interpreted as a (total) derivative in the strict sense. However, in a rather loose sense, it may be interpreted as something like a total derivative. It tells us by how much the aggregate output Y changes when an additional marginal unit of resource material R_i is used in production, by dividing it up among the m intermediate good production processes in such a manner that $dR_i = \sum_{j=1}^{m} dR_{ij}$.[9]

Again, for all materials $i \in I$ which make up the final good, ρ_i is strictly positive so that, according to Inequality (3.27), $1/\rho_i < +\infty$ is a a finite upper bound on dF/dR_i. From Assumption 3.1 it follows, that there is at least one such material. For all other materials with $i \notin I$, ρ_i is zero so that $1/\rho_i$ is not a finite upper bound. This establishes the following result.

Proposition 3.4

(i) For all materials $i \in I$, which make up the final good, the marginal product of resource material i in aggregate production, dF/dR_i, is bounded from

---

[9]In that sense, one could define

$$\frac{dF}{dR_i} \equiv \sum_{j=1}^{m} \frac{\partial F}{\partial Y_j} \frac{\partial F^j}{\partial R_{ij}} \quad \text{subject to} \quad dR_i = \sum_{j=1}^{m} dR_{ij} \; .$$

However, this definition is not unique. While there is a multitude of ways in which dR_i may be divided up among the m different intermediate good production processes, Inequality (3.26) holds in any case. Hence, result (3.27) holds irrespective of the exact way in which dF/dR_i is defined.

above by the inverse of this material's mass fraction in the final good, $1/\rho_i$.

(ii) There exists at least one such material for which the marginal product is bounded from above.

(iii) For all materials $i \notin I$*, which are not contained in the final good, the marginal product of resource material* i *in aggregate production,* dF/dR_i*, does not need to be bounded from above.*

Comparing Propositions 3.3 and 3.4 for macro level production with Propositions 3.1 and 3.2 for micro level production, we see that all results that were obtained in the simple micro-level setting essentially carry over to the general two-level-multi-resources-multi-processes setting. The only qualification is that the boundedness-results only hold for materials in the set I which make up the final good.

3.5 Discussion

It has been shown that the Inada conditions, when applied to material resource inputs, may be inconsistent with the thermodynamic law of conservation of mass, the so-called Materials-Balance-Principle. In particular, the analysis has revealed that the average and marginal product of a natural resource material in aggregate production are bounded from above due to the thermodynamic law of mass conservation if the final good, an all-purpose commodity, contains this material. An upper bound is given by the inverse of this material's mass fraction in the final good.

The analysis was based on a model of multi-level production where different intermediate goods are produced from different elementary resources, and an all-purpose final commodity is produced from these intermediates. Note that no limits on substitution between resource materials or between intermediate products have been assumed. Another thing to note is that the upper bound specified by inequalities (3.18) and (3.27) is certainly not the lowest upper bound, but comes out of a more or less crude estimation (from Equations (3.17), (3.25) to Inequalities (3.18), (3.26)). For this reason, the upper bound given here does not depend on any model parameters other than ρ_i.

When discussing the relevance of these results for the natural-resources-and-economic-growth-debate, the crucial questions are:

(i) How many, and which natural resource materials are elements of the set I? That is, what are the natural resource materials that make up, materially, the final good?

(ii) What is these materials' mass fraction ρ_i in the final good?

(iii) How do the set I and the relevant parameter values ρ_i change over time?

It is probably the difference in opinion on these questions which make a difference between the 'optimists' and the 'pessimists' in the discussion about the thermodynamic limits to economic growth.

This analysis has revealed that there are stringent thermodynamic limits to resource productivity in aggregate production for a number of natural resource materials. The analysis has also revealed that not all resource materials need to be limiting. Hence, the overall conclusion is that the question of thermodynamic limits to economic growth requires a detailed investigation, with separate analyses and results for each material. This shifts the focus of the debate from overall growth to the more detailed level of factor substitution and structural change.

4. Temporal and Thermodynamic Irreversibility in Production Theory*

4.1 Introduction

From a physical point of view, irreversibility is an essential dynamic feature of real production. Therefore, it should be properly taken into account in dynamic analyses of production systems.[1] For example, engineers account for irreversibility when designing and optimizing production processes (Bejan 1997, Bejan et al. 1996, Brodyansky et al. 1994, Szargut et al. 1988), and economists consider irreversibility when studying economy-environment interactions (Ayres 1998, 1999b, Baumgärtner et al. 2006, Faber et al. 1995[1983], Georgescu-Roegen 1971, Mäler 1974, Perrings 1987, Pethig 1979).

The idea of irreversibility can be rigorously rooted in the laws of nature (Zeh 2001), in particular in thermodynamics (Kondepudi and Prigogine 1998: 84ff). The importance of thermodynamic irreversibility, and the physicists' preoccupation with this concept, lies in the fact that it precludes the existence of perpetual motion machines, that is, devices which use a limited reservoir of available energy to perform work forever (Second Law of Thermodynamics). It is an everyday experience that no such thing as a perpetual motion machine exists. In order to make this insight accessible to economic analysis, and to the study of long term economy-environment interactions, it is necessary to adequately represent thermodynamic irreversibility as a constraint for economic action (Georgescu-Roegen 1971).[2]

*Sections 4.2 and 4.3.2 of this Chapter have previously been published in *Economic Theory*, **26**(3), 725–728 (2005). They formally elaborate an idea which has originally been proposed by Baumgärtner (2000: Section 11.1).

[1]In this chapter, the focus is on irreversibility in processes of production. Other origins of irreversibility, for instance investment and learning under uncertainty (e.g. Dixit 1992, Dixit and Pindyck 1994, Pindyck 1991) or 'lock-in' due to increasing returns (Arthur 1989), are not considered here. For a comprehensive survey of various notions of irreversibility in economics see Dosi and Metcalfe (1991).

[2]The contribution of Georgescu-Roegen (1971), who considered the Second Law of Thermodynamics 'the most economic of all physical laws' (p. 280), initiated a heated debate over

Economists have devoted some effort to incorporating irreversibility into production theory. The reason is primarily a concern for physical realism in the description of the set of 'feasible' production processes.[3] However, irreversibility has often been introduced into the theory as an ad-hoc-assumption. As a result, the assumption did not always achieve what it actually should achieve from a thermodynamic point of view, namely to imply irreversibility of the system's evolution as stated by the Second Law of Thermodynamics.

In this chapter, I will introduce a formal and rigorous definition of thermodynamic irreversibility, which is (i) sound from a physical point of view and (ii) formulated such that it is compatible with formal modelling in economic production theory. In order to assess, whether – and to what extent – different notions of irreversibility from production theory capture thermodynamic irreversibility, I will then reexamine two prominent irreversibility concepts – the one due to Koopmans (1951b) and the one due to Arrow-Debreu (Arrow and Debreu 1954, Debreu 1959) – against the definition of thermodynamic irreversibility.

The chapter is organized as follows. In Section 4.2, I briefly review the concept of *thermodynamic irreversibility* and distinguish it from the weaker concept of *temporal irreversibility*. I propose formal definitions of both concepts. In Section 4.3, the irreversibility concepts of Koopmans and of Arrow-Debreu are reexamined against these definitions. I show that Koopmans' notion of irreversibility fully captures thermodynamic irreversibility, and that the notion of Arrow-Debreu does not capture thermodynamic irreversibility but only the weaker aspect of temporal irreversibility. I conclude with Section 4.4, by putting the results into perspective.

4.2 The Thermodynamic Notion of Irreversibility

The textbook definition of thermodynamic irreversibility builds on the consideration of an isolated system, defined by its boundaries.[4] A transformation

---

the question of whether the Entropy Law is relevant for economics (e.g. Burness et al. 1980, Daly 1992b, Kåberger and Månsson 2001, Khalil 1990, Lozada 1991, 1995, Norgaard 1986, Townsend 1992, Williamson 1993, Young 1991, 1994). See Baumgärtner et al. (1996) for a summary of this discussion.

[3]In economic theory, irreversibility of the total production set has originally been considered as fundamental (beside closedness, convexity and the no-free-lunch condition) for the existence proof of general competitive equilibrium in economies with production (Arrow and Debreu 1954, Debreu 1959). Yet, later it has become obvious that, while this assumption simplifies the proof, it is not necessary (Debreu 1962: 258; Koopmans 1957: 78, Footnote 4; McKenzie 1959: 55, 1961).

[4]Recall (from Section A2.1) that a thermodynamic system is called *isolated* if it exchanges neither energy nor matter with its surrounding environment; it is called *closed* if it exchanges

over time of an isolated system between some initial state and some final state is called *irreversible* if there is no means by which the system can be exactly restored to its initial state (Kondepudi and Prigogine 1998, Zeh 2001). Otherwise, the transformation is called *reversible*.

As for closed or open systems, one can consider the system and its environment, such that the overall system is once again isolated. Thus, a transformation in an open system is called irreversible if there is no means by which the system and its environment can be exactly restored to their respective initial states. Since the economy is an open system in the thermodynamic sense, any physically meaningful analysis of economic irreversibility should consider the economy plus its natural environment ('Planet Earth').

Formalizing this notion of thermodynamic irreversibility in a way customary to economic theory requires one

(i) to distinguish between *states* of the system (stock variables) and *transformations* of the system (flow variables), and

(ii) to consider time as an explicit variable.

Consider an economy with n physically different goods, including natural resources and wastes, and T discrete points in time. Let $s_i(t) \in \mathbb{R}_+$ denote the stock of good i $(i = 1, \ldots, n)$ at time $t \in [1, \ldots, T]$ and $s(t) = (s_1(t), \ldots, s_n(t)) \in \mathbb{R}^n_+$. At every point in time, s completely characterizes the state of the economy in terms of the different state variables s_i. This is an *explicit time representation (ETR)* of the commodity space, since time shows up explicitly. One could also adopt an *implicit time representation (ITR)*, by making the following assumption (Arrow and Debreu 1954: 266, Debreu 1959: 29).

Assumption 4.1
The same physical commodity at two different points in time is regarded as two different economic commodities. (Dated goods)

In order to compare different notions of irreversibility it is helpful to have a mapping from ITR to ETR-representations. Let $Y \subseteq \mathbb{R}^{nT}$ be the set of all feasible aggregate ITR-production vectors y, that is, the set of all feasible transformations of the system. A production vector y has as components the net output of all dated commodities. For the sake of notational convenience assume that all inputs enter a production process y simultaneously at time $\underline{t}(y)$ and that all outputs are simultaneously obtained at $\overline{t}(y)$.[5] One may then define the mapping

$$\Pi : \mathbb{R}^{nT} \to \mathbb{R}^{n+2} \qquad \text{with} \qquad \Pi(y) = (\hat{y}, \underline{t}(y), \overline{t}(y)) \ , \tag{4.1}$$

energy, but not matter; and it is called *open* if it exchanges both energy and matter with its surrounding environment.

[5]In other words, consider *elementary* production processes (Takayama 1985: 487–488).

where $\hat{y} \in \mathbb{R}^n$ is the vector of physical net output and $\underline{t}$ ($\overline{t}$) denotes the point in time at which inputs (outputs) are supplied (obtained) under transformation y. For every ITR-production vector $y \in Y$ the image $\Pi(y)$ is the corresponding ETR-production vector. With this notation, the effect of a transformation y on the state s of the system is given in ETR-terminology by

$$s(\overline{t}(y)) = s(\underline{t}(y)) + \hat{y} \ . \tag{4.2}$$

In order to assess the dynamic effect of several transformations on the state of the system, one needs to make an assumption about what combinations of transformations are feasible. For that sake, I will assume throughout this chapter that the ITR-production set Y has the following property:

Assumption 4.2
If $y^1, y^2 \in Y$ and $\underline{t}(y^j) \geq \overline{t}(y^i)$ for $i, j = 1, 2$ and $i \neq j$, then $y^1 + y^2 \in Y$. (Temporal additivity)

In words, if both y^1 and y^2 are feasible ITR-production vectors and one of them (y^j) begins after the other one (y^i) has ended, then it is feasible to carry out first the physical transformation described by y^j and then, later in time, the physical transformation described by y^i. The property is called *temporal* additivity since it refers solely to adding two production processes in the time dimension. This assumption is considerably weaker than the usual assumption of additivity ($y^1, y^2 \in Y$ implies $y^1 + y^2 \in Y$) since only the addition of production vectors which are carried out *one after the other* is assumed to be feasible. In contrast, ordinary additivity would also allow for the addition of *simultaneous* physical transformations, thus ruling out decreasing returns to scale.[6] One can now define thermodynamic irreversibility as follows.

Definition 4.1
An ITR-production set Y has the property of *thermodynamic irreversibility* if and only if for every $y \in Y$ with $\hat{y} \neq 0$ there exists no $y' \in Y$ with $\hat{y}' = -\hat{y}$.

In words, if y is a feasible non-trivial production vector there exists no feasible production vector which reverses the physical net effect of y. Hence,

---

[6]While temporal additivity (Assumption 4.2) is considerably weaker than full additivity ($y^1, y^2 \in Y$ implies $y^1 + y^2 \in Y$) one may, however, still imagine examples where this assumption is not fulfilled. For instance, when the change in the state of the system induced by the first transformation makes it impossible to carry out the second transformation later on. The following system would be a (disgusting!) example: Transformation y^1 uses one cup of black coffee and a drop of vanilla syrup at time t_0 to produce vanilla flavored coffee at time $t_1 > t_0$; transformation y^2 uses one cup of black coffee and one drop of banana syrup at time t_2 to produce banana flavored coffee at time $t_3 > t_2$. If the system has an initial stock of only one cup of black coffee at time t_0, both transformations y^1 and y^2 are feasible separately, but it is not feasible to carry out both of them. What Assumption 4.2 essentially states is that *stocks* are not constraining the *transformations* of the system.

under thermodynamic irreversibility the initial state of the system cannot be restored: if the system has initially been in some state s^0, and as a consequence of the transformation y has evolved into some final state $s^f = s^0 + \hat{y}$, then there is no transformation y' which brings the system into a state which is identical to the initial state s^0.[7]

Note that the crucial condition ($\hat{y}' = -\hat{y}$) in Definition 4.1 is in terms of the physical net effect of transformations y and y'. It does not constrain in any way the time structure of these transformations, that is, when inputs are supplied and outputs are obtained. A weaker restriction on the set of feasible production vectors than thermodynamic irreversibility is temporal irreversibility.

Definition 4.2
An ITR-production set Y has the property of *temporal irreversibility* if and only if for every $y \in Y$ with $\hat{y} \neq 0$ there exists no $y' \in Y$ with $\hat{y}' = -\hat{y}$ and $\underline{t}(y') = \overline{t}(y)$, $\overline{t}(y') = \underline{t}(y)$.

In words, if y is a feasible non-trivial production vector, there exists no y' which reverses both the physical net effect of transformation y and its temporal input-output-structure.

While thermodynamic irreversibility excludes the possibility that the system returns into its initial state s^0 at any, possibly later, point in time $t \geq t_0$, temporal irreversibility only excludes the possibility that the system returns into its initial state s^0 at initial time t_0. Temporal irreversibility, thus, states that one cannot reverse physical transformations by going back in time. Obviously, temporal irreversibility is a weaker concept than thermodynamic irreversibility, in the sense that thermodynamic irreversibility implies temporal irreversibility, but not vice versa.[8]

Proposition 4.1
If an ITR-production set Y has the property of thermodynamic irreversibility, it also has the property of temporal irreversibility.

4.3 Notions of Irreversibility in Production Theory

Having now a formal and rigorous definition of thermodynamic irreversibility (Definition 4.1), and also of the weaker concept of temporal irreversibility

---

[7]The condition $s^f + \hat{y}' = s^0 + \hat{y} + \hat{y}' = s^0$ can only be fulfilled for $\hat{y}' = -\hat{y}$, which is precluded by thermodynamic irreversibility according to Definition 4.1.

[8]The condition on Y in Definition 4.2 of temporal irreversibility includes the condition in Definition 4.1 of thermodynamic irreversibility ($\hat{y}' = -\hat{y}$), and puts additional conditions on Y ($\underline{t}(y') = \overline{t}(y)$, $\overline{t}(y') = \underline{t}(y)$).

(Definition 4.2), one can now reexamine the irreversibility concept of Koopmans (1951b) and the one of Arrow-Debreu (Arrow and Debreu 1954, Debreu 1959) against these definitions.[9]

4.3.1 Koopmans' Notion of Irreversibility

Koopmans' (1951b) theory of production is based on the analysis of activities, that is, vectors representing feasible combinations of inputs and outputs. His analysis is static. In the language introduced in Section 4.2 above, all of Koopmans' statements refer to the physical net effect $\hat{y}$ of transformations y. There are two basic assumptions associated with the notion of an activity: *divisibility* and *additivity* (Koopmans 1951b: 36).

Assumption 4.3
Each activity is capable of continuous proportional expansion or reduction. (Divisibility)

Assumption 4.4
Any number of activities can be carried out simultaneously without modification in the technical rations by which they are defined, provided only that the total resulting net output of any commodity, whenever negative, is within the limitations on primary resources. The joint net output of any commodity from all activities then equals the sum of the net outputs of that commodity from the individual activities. (Additivity)

Assumptions 4.3 and 4.4 imply that all statements about feasible activities also hold for linear combinations of feasible activities. Thus, they exclude economies or diseconomies of scale. Note that Koopmans' additivity assumption (Assumption 4.4) is considerably stronger then the assumption of temporal additivity (Assumption 4.2), since it refers to the *simultaneous* addition of physical transformations.

Koopmans (1951b: 48) then introduces the idea of irreversibility in the form of a fundamental postulate:

Postulate A
It is impossible to find a set of positive amounts of some or all activities, of which the joint effect is a zero net output for all commodities.

This is to say (Koopmans 1951b: 48–49) that it is not possible to find activity vectors, such that the net output resulting from one of them is being exactly

---

[9]The treatment of Arrow-Debreu has become the state-of-the-art way of incorporating irreversibility into production theory. It is essentially what is taught in many economic textbooks (e.g. Mas-Colell et al. 1995: 132).

offset by the net output brought about by a linear combination of the other activities. Koopmans' Postulate A thus excludes the possibility that by a suitably chosen combination of activity vectors the system undergoes some activites as a result of which it returns back to its initial state. Obviously, Koopmans' notion of irreversibility is exactly one of thermodynamic irreversibility as specified by Definition 4.1 above.[10]

4.3.2 The Arrow-Debreu Notion of Irreversibility

A formalized and slightly, but significantly, altered version of Koopmans' irreversibility concept is introduced by Arrow and Debreu (1954) as well as Debreu (1959). Their irreversibility concept consists of two elements:

(i) a formal statement about the set of feasible production vectors and

(ii) the ITR-convention of dated goods (Assumption 4.1)

The formal statement (i) is the following (Arrow and Debreu 1954: 267, Debreu 1959: 40):

Assumption 4.5
$Y \cap -Y \subseteq \{0\}$. (Arrow-Debreu-Irreversibility)

This can be reformulated as saying that if a production vector y is feasible, the reverse production vector $-y$ is not feasible unless y describes null-activity ($y = 0$). While this is, as it stands, just a rephrasement of Koopmans' Postulate A in more technical language, the ITR-convention (ii) constitutes a new element for the notion of irreversibility.

To illustrate how the dated-goods-interpretation (Assumption 4.1) affects the working of Assumption 4.5, consider an economy with two physically distinct commodities, say metal and screws, and two points in time, t_0 and $t_1 > t_0$. Initially, the state of the system is

$$s(t_0) = (M, S) \ , \tag{4.3}$$

with M denoting the total initial stock of metal in the economy and S the total initial stock of screws. Assume that there exists a feasible production process y which turns $m > 0$ units of metal at time t_0 into one unit of screws at time t_1. The corresponding ETR-production vector reads $\Pi(y) = ((-m, +1), t_0, t_1)$. This transformation changes the state of the system to

$$\begin{aligned} s(t_1) &= s(t_0) + \hat{y} \\ &= (M, S) + (-m, +1) \\ &= (M - m, S + 1) \ . \end{aligned} \tag{4.4}$$

[10]Note that Tjalling C. Koopmans by training was a physicist and his first two publications were in physics (Niehans 1990: 408).

By Assumption 4.1, there are four ITR-goods: metal at time t_0, metal at time t_1, screws at time t_0 and screws at time t_1. With the convention that production vectors have as components net output of metal at time t_0, net output of metal at time t_1, net output of screws at time t_0, and net output of screws at time t_1, the production process y is represented by the ITR-production vector $y = (-m, 0, 0, +1)$. By Assumption 4.5, one has $-y = (+m, 0, 0, -1) \notin Y$. This means, it is not possible to turn one unit of screws at time t_1 back into m units of metal at time $t_0 < t_1$, thus bringing back the system to its original state $s(t_0) = (M, S)$ at the initial point in time t_0 (Figure 4.1).

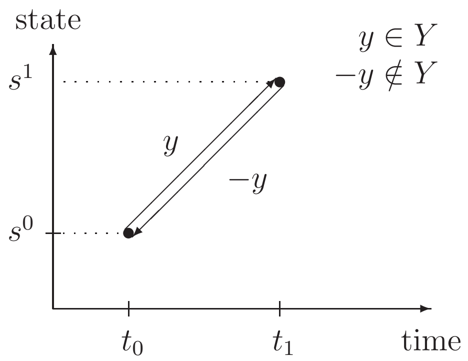

Figure 4.1 The Arrow-Debreu notion of irreversibility establishes temporal irreversibility. (Figure from Baumgärtner 2000: 236).

Obviously, Assumptions 4.1 and 4.5 establish temporal irreversibility. But they do not suffice to establish thermodynamic irreversibility.

Proposition 4.2

(i) Assumptions 4.1 and 4.5 imply that Y has the property of temporal irreversibility.

(ii) Assumptions 4.1 and 4.5 do not imply that Y has the property of thermodynamic irreversibility.

Proof: (i) is obvious. (ii) is proven by giving an example for an ITR-production set Y which satisfies Assumptions 4.1 and 4.5, but does not have the property of thermodynamic irreversibility.

Consider the metal-and-screws-economy introduced above with four points in time, $t_0 < t_1 < t_2 < t_3$. Assume that there are two feasible production processes: the first one (y^1) turns $m > 0$ units of metal at time t_0 into one unit of screws at time t_1; the other one (y^2) turns one unit of screws at time t_2 into

m units of metal at time t_2. The corresponding ETR-production vectors read $\Pi(y^1) = ((-m, +1), t_0, t_1)$ and $\Pi(y^2) = ((+m, -1), t_2, t_3)$.

Under Assumption 4.1, there are eight distinct ITR-goods: metal at time t_0, metal at time t_1, metal at time t_2, metal at time t_3, screws at time t_0, screws at time t_1, screws at time t_2 and screws at time t_3. With the convention that the components of the production vectors represent net output of metal at times t_0, t_1, t_2, t_3, and net output of screws at time t_0, t_1, t_2, t_3, the production possibilities can be represented by the ITR-production set $Y = \{y^1, y^2\}$ with $y^1 = (-m, 0, 0, 0, 0, +1, 0, 0)$ and $y^2 = (0, 0, 0, +m, 0, 0, -1, 0)$.

Assumption 4.5 is satisfied, as $y^2 \neq -y^1$. But with $\hat{y}^1 = (-m, +1)$ and $\hat{y}^2 = (+m, -1)$ one has $\hat{y}^2 = -\hat{y}^1$, in contradiction of thermodynamic irreversibility. □

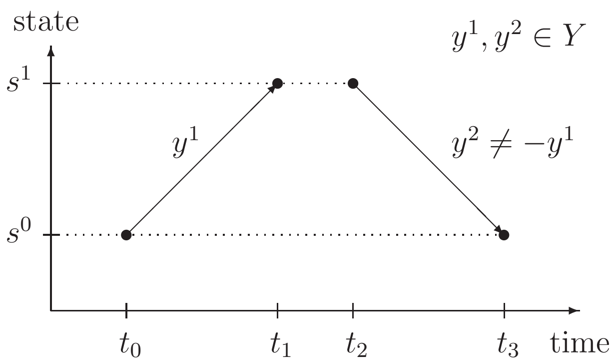

Figure 4.2 The Arrow-Debreu notion of irreversibility does not establish thermodynamic irreversibility. (Figure modified from Baumgärtner 2000: 237).

The example used in the proof can be interpreted as follows (Figure 4.2). Initially, the state of the system is again

$$s(t_0) = (M, S) \; , \tag{4.5}$$

with M denoting the total initial stock of metal in the economy and S the total initial stock of screws. As a consequence of carrying out production process y^1, which turns $m > 0$ units of metal at time t_0 into one unit of screws at time t_1, the state of the economy changes from its initial state $s(t_0)$ to

$$\begin{aligned} s(t_1) &= s(t_0) + \hat{y}^1 \\ &= (M, S) + (-m, +1) \\ &= (M - m, S + 1) \; . \end{aligned} \tag{4.6}$$

That is, the stock of metal has decreased by m units, and the stock of screws has increased by one unit. If then production process y^2 is carried out, which is also feasible and turns one unit of screws at time t_2 into m units of metal at time t_2, the state of the economy changes from $s(t_2) = s(t_1)$ to

$$\begin{aligned} s(t_3) &= s(t_2) + \hat{y}^2 \\ &= (M - m, S + 1) + (+m, -1) \\ &= (M, S) \ , \end{aligned} \tag{4.7}$$

thus restoring the initial state of the system: $s(t_3) = s(t_0)$. Hence, by carrying out first y^1 and subsequently y^2 the economy would undergo a cyclical transformation process: some amount of metal is turned into screws, and later all screws are transformed back into the original amount of metal. This simple model would be the blueprint for a perpetual motion machine, in contradiction of thermodynamic irreversibility

4.4 Conclusion

Summing up, while Koopmans' notion of irreversibility fully captures thermodynamic irreversibility, the Arrow-Debreu notion does not capture thermodynamic irreversibility but only the weaker aspect of temporal irreversibility. The crucial difference between the two concepts is Arrow-Debreu's interpretation of goods as being dated (Assumption 4.1). While the dated-goods-interpretation allows one to obtain a theory of time and uncertainty with seemingly no formal effort (Debreu 1959: 98), it considerably reduces the physical content of the irreversibility assumption proper (that is, Assumption 4.5).[11,12] Had Arrow-

[11] It is interesting to note that Debreu uses the dated-goods-interpretation only when dealing with uncertainty (Debreu 1959: Chapter 7). Yet, for this sake he considers an exchange economy without production (or irreversibility). On the other hand, when studying production (and irreversibility) he does not make any use of the dated-goods-interpretation.

[12] The usefulness of introducing time into production theory by considering goods as being dated may be questioned anyway (Baumgärtner 2000: 239). If one commodity (as described by its physical properties) at one given point in time and the very same commodity (as described by its physical properties) at another point in time are taken as two different economic goods, then such an economic description creates artificial processes of production where actually no physical transformation of energy and matter takes place. For example, simply letting an object rest in some place for a while would constitute a process of production under such an interpretation, since one economic good – the object at an earlier time – is transformed into a different economic good – the very same object at some later time. In this view, economic production may happen where actually no physical transformation of energy and matter takes place and no positive amount of entropy is created. So, the dated-goods-interpretation creates a notion of production which is at odds with the physical view of production as a transformation of energy and matter.

Debreu not made the dated-goods-interpretation, their notion of irreversibility would be fully equivalent to Koopmans' one.

In order to put these results into perspective, one should add two comments. First, Arrow and Debreu were not primarily concerned with making realistic assumptions, but with identifying the weakest, and thus most general, assumptions under which a general competitive equilibrium could be shown to exist.[13] And indeed, temporal irreversibility is a weaker assumption than thermodynamic irreversibility (Proposition 4.1). As a consequence, all results obtained in an Arrow-Debreu-framework, including their assumption of irreversibility, apply as well to systems for which the more restrictive assumption of thermodynamic irreversibility is made.

Second, while thermodynamic irreversibility is a fact of nature for completely specified thermodynamic systems – that is, isolated, closed or open systems which are described in terms of *all* state variables – it does not need to hold in an incompletely specified system (Dyckhoff 1994: 78). For example, turning metal first into screws and then completely back into the original amount of metal, may appear feasible as long as one neglects energy. But transforming metal into screws requires energy, thus reducing the stock of available energy in the system; and so does the recycling of metal from screws, which further decreases the stock of available energy in the system. Therefore, the transformation of metal into screws, which seems to be reversible when neglecting the state variable 'available energy', turns out to be actually irreversible when properly taking all physical state variables into account.[14]

In the end, the relevance of thermodynamic irreversibility for economic analysis comes down to the question of which system is under study. Arrow and Debreu's description of production vectors includes 'only the components which correspond to marketable commodities' (Arrow and Debreu 1954: 267). This system is incompletely specified from the physical point of view, because the commodity space does not include essential physical state variables such as available energy or entropy. Therefore, there is no reason to take thermodynamic irreversibility to be a relevant property of this system. But if one aims at an encompassing analysis of economy-environment interactions, essential physical state variables have to be included in the description of the system. Thermodynamic irreversibility then is a relevant property of the system, and the Arrow-Debreu notion of irreversibility is too weak to be in full accordance with the laws of nature.

---

[13] But see Footnote 3 above.

[14] Baumgärtner (2000: Chapter 11) further elaborates on the 'complete-representation'-approach to modeling irreversibility, based on the concept of joint production.

5. Necessity and Inefficiency in the Generation of Waste*

with Jakob de Swaan Arons

5.1 Introduction

The sheer amount of waste generated in modern industrial economies is enormous. For example, in 1990 the amount of waste in West Germany (measured in physical units, such as tons) exceeded the amount of useful economic output (also measured in physical units) by more than a factor of four: out of a total material output of 59,474.6 million metric tons generated by all sectors of the economy, only 3,602.6 million metric tons (6.1 %) were contained in the different components of GDP, while 7,577.2 million metric tons (12.7 %) were intermediate outputs for reuse within the economy (including recovered and recycled materials) and 48,294.8 million metric tons (81.2 %) were final wastes (Statistisches Bundesamt 1997). This huge dimension of material waste generation is also confirmed for other industrialized countries, e.g. Denmark, Italy and the USA (Acosta 2001).

The notion of 'waste' is a difficult one, as a proper definition should build on descriptive materials-balance on the one hand, and normative human attitudes and valuation on the other hand (Bisson and Proops 2002). 'Waste' essentially denotes an ultimately unwanted by-product in the production of some desired good or service. We will use the term here in a slightly more general sense, to refer to a by-product of a desired good on the level of a single production process. For example, the process of enriching uranium generates depleted uranium as a by-product together with the desired enriched uranium. Although this by-product may be used to produce other products (e.g. special ammunition), it is a 'waste' in the process of enriching uranium. Thus, we focus on the material origin of waste. At the same time, we disregard two related issues which are most relevant in the discussion of waste, and which

*Previously published, without the Appendix, in *Journal of Industrial Ecology*, **7**(2), 113–123 (2003). The Appendix is a revised version of Baumgärtner (2000: 72–77).

are treated in detail elsewhere in the literature. First, we do not explicitly analyze whether the by-product ('waste') considered here is actually positively or negatively valued, as this would require an economic analysis (Baumgärtner 2000, 2004d, Powell et al. 2002). Instead, we argue that in many cases one can safely assume that it is unwanted. Second, by looking only at by-products on the level of a production process, we do not follow the by-products' broader impact in an economy. In particular, we do not consider the possibility that what is an unwanted by-product for one producer may be a valuable resource for another producer, giving rise to the idea of an 'industrial ecology' (Ayres and Ayres 2002, Hardy and Graedel 2002). With such a notion of 'waste', our analysis is relevant for the field of industrial ecology, because it deals with the qualitative and quantitative conditions under which all those 'waste' by-products come into existence that give rise to the problems commonly studied in the industrial ecology literature.

It has been argued, based on the thermodynamic laws of mass conservation and entropy generation, that in industrial production processes the occurrence of waste is as necessary as the use of material resources (Ayres and Kneese 1969, Faber et al. 1998, Georgescu-Roegen 1971).[1] On the other hand, it seems to be quite obvious that the sheer amount of waste currently generated is to some extent due to various inefficiencies and might, in principle, be reduced.

In this chapter we analyze the question to what extent the occurrence of waste is actually an unavoidable necessity of industrial production, and to what extent it is an inefficiency that may, in principle, be reduced. For that sake, we employ the laws of thermodynamics as an analytical framework within which results about current 'industrial metabolism' (Ayres and Simonis 1994) may be rigorously deduced in energetic and material terms.

In Section 5.2, we demonstrate that industrial production is necessarily and unavoidably joint production. This means waste outputs are an unavoidable by-product in the industrial production of desired goods. In Section 5.3, we analyze the degree of thermodynamic (in)efficiency of industrial production processes, and the associated amounts of waste due to these inefficiencies. Section 5.4 concludes.

5.2 Joint Production of Desired Goods and Waste

5.2.1 The Thermodynamic View of Production

Production can in the most general way be conceived of as the transformation of a number of inputs into a number of outputs. In thermodynamic terms, en-

[1]For example, Georgescu-Roegen (1975: 357) has argued that 'waste is an output just as unavoidable as the use of natural resources'.

ergy (actually: exergy) and matter are the fundamental factors of production (Ayres 1998, Baumgärtner 2000, Faber et al. 1998, Ruth 1993). From a thermodynamic view,[2] two quantifiable characteristics of an input or an output are its mass, m, and its entropy, S. Alternatively, one could use its exergy instead of its entropy; this will be done in the next section. Because both mass and entropy are extensive quantities, it is useful to introduce the ratio of the two, $\sigma = S/m$, for $m > 0$ as an intensive quantity. σ is called *specific entropy* and measures the entropy per unit mass of a bulk of matter irrespective of that bulk's size.[3]

5.2.2 Joint Products are Unavoidable in Industrial Production

Let us now narrow down the analysis to the particular type of production which is found in most developed countries and which is most relevant as far as economy-environment interactions are concerned. This is what one may call *regular industrial production.* For that sake, consider the following reference model of regular industrial production (Baumgärtner 2000: Chapter 4). A raw material is transformed into a final product. The exergy necessary to carry out that transformation is typically provided by a material fuel. As the analysis of the reference model will reveal, it is then unavoidable that a by-product is jointly produced with the desired product. The analysis will also suggest that this by-product may often be considered an unwanted waste. The industrial production process can, thus, be depicted as in Figure 5.1. An example of such an industrial production process is the production of pure iron as a desired product from iron ore as raw material (see e.g. Ruth 1995a). The fuel in that example is coke, and there are slag, carbon dioxide and heat as waste by-products.

The focus on regular industrial production processes justifies building the reference model on the assumption of two kinds of inputs, raw material and fuel, and not more than two kinds of outputs, desired product and by-product.[4]

---

[2]For those readers not familiar with classical thermodynamics we recommend the work of Callen (1985), Kondepudi and Prigogine (1998) or Zemansky and Dittman (1997) as comprehensive, yet accessible introductions. The appendix to Chapter 2 provides a short and basic introduction to classical thermodynamics in non-technical terms. Bejan (1997) gives a good introduction to engineering thermodynamics.

[3]Thermodynamic variables, such as volume and particle number, which are proportional to the size of the system, are called *extensive* variables. Variables, such as temperature or pressure, that specify a local property and are independent of the size of the system, are called *intensive* variables. If one doubled a bulk of matter, then the two extensive quantities m and S would double as well while the ratio of the two, $\sigma = S/m$, would remain constant. Specific entropy, thus, is an intensive variable.

[4]This assumption may be relaxed. It may be assumed that there are a number of additional inputs and outputs besides the ones mentioned in the text. The joint production result is not altered by the assumption of additional inputs or outputs.

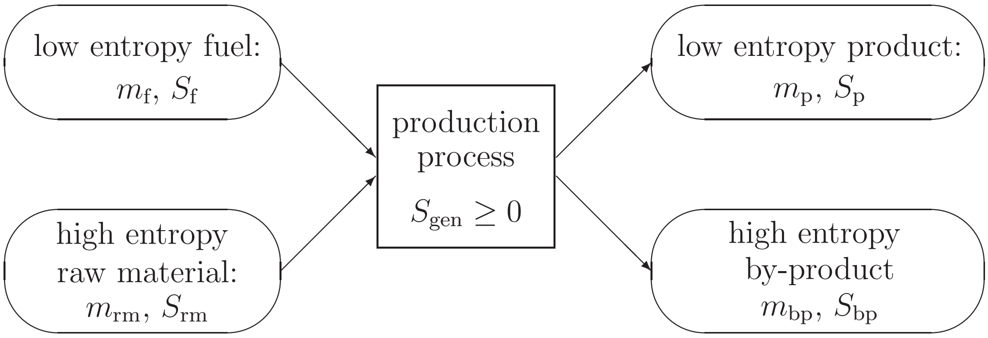

Figure 5.1 The thermodynamic structure of regular industrial production in terms of mass and (specific) entropy. (Figure modified from Baumgärtner 2000: 74).

In the notation introduced above, m_j and S_j are the mass and the entropy of the inputs and outputs involved and σ_j is their respective specific entropy ($j = \mathrm{rm}, \mathrm{f}, \mathrm{p}, \mathrm{bp}$ which stands for raw material, fuel, product, by-product). One may then formally define the notion of industrial production in thermodynamic terms.

Definition 5.1
Within the formal framework of the reference model, a process of production is called *industrial production* if and only if it exhibits the following three properties:

$$m_\mathrm{rm}, m_\mathrm{p} > 0, \tag{5.1}$$
$$\sigma_\mathrm{rm} > \sigma_\mathrm{p}, \tag{5.2}$$
$$m_\mathrm{f} > 0. \tag{5.3}$$

Property (5.1) means that the production process essentially consists of a material transformation, that is, a raw material is transformed into a material desired product. Property (5.2) states that the direction of this material transformation is such as to transform a raw material of relatively high specific entropy into a desired product of lower specific entropy. In our example, iron oxide (Fe_2O_3) and pure iron (Fe) have a specific entropy of 87.4 J/mole K and 27.3 J/mole K respectively (see Table 5.1; Kondepudi and Prigogine 1998. Appendix). The underlying idea is that most raw materials are still impure and, therefore, can be thought of as mixtures from which the desired product is to be obtained by de-mixing of the different components of the raw material. More generally, desired products are thought of as matter in a more orderly state than the raw material. From basic thermodynamics we know that such a transformation process requires the use of exergy. Property (5.3) states that the exergy input also has mass, that is, the exergy necessary to carry out the desired transformation is provided by a material fuel, such as for example, oil, coal or gas.

The constraints imposed on production processes by the laws of thermodynamics can be formalized as follows:

$$m_{\mathrm{rm}} + m_{\mathrm{f}} = m_{\mathrm{p}} + m_{\mathrm{bp}}, \tag{5.4}$$

$$S_{\mathrm{rm}} + S_{\mathrm{f}} + S_{\mathrm{gen}} = S_{\mathrm{p}} + S_{\mathrm{bp}} \quad \text{with} \quad S_{\mathrm{gen}} \geq 0. \tag{5.5}$$

The Law of Mass Conservation (Equation 5.4) states that the total ingoing mass has to equal the total outgoing mass because mass is conserved in the production process. The Second Law of Thermodynamics (Equation 5.5), the so-called Entropy Law, states that in the production process a non-negative amount of entropy is generated, S_{gen}, which is added to the total entropy of all inputs to yield the total entropy of all outputs.

Within the framework of that reference model, the two laws of thermodynamics, Equations (5.4) and (5.5), together with the assumption of industrial production, Properties (5.1)–(5.3), imply that the second output necessarily exists (Baumgärtner 2000: 77).

Proposition 5.1
For any process of industrial production of a desired product (Properties 5.1–5.3), the laws of thermodynamics (Equations 5.4 and 5.5) imply the existence of at least one by-product.

Proof: see Appendix A5.1.

This means, the occurrence of a by-product is necessary and unavoidable in every process of regular industrial production. In economic terms, one may speak of 'joint production', as the desired product and the by-product are necessarily produced together (Baumgärtner et al. 2001, 2006).

The intuition behind this result is the following. One obvious reason for the existence of joint outputs besides the desired product is simply conservation of mass. If, for instance, pure iron is produced from iron ore with a carbon fuel, the desired product, which is pure iron, does not contain any carbon. Yet, the carbon material from the fuel has to go somewhere. Hence, there has to be a joint product containing the carbon. But there is a second reason for the existence of joint products besides and beyond conservation of mass, and that is the generation of entropy according to the Second Law of Thermodynamics. Think of a production process where all of the raw material and the material fuel end up as part of the desired product, for example, the production of cement. In that case, mass conservation alone would not require any joint product. But because the desired product has lower specific entropy than the raw material, and there is some non-negative amount of entropy generated by the process, there is a need for a joint output taking up the excess entropy. In many cases, as in the example of cement production, this happens in the form of low-temperature heat, which may be contained in the product, a by-product or transferred to the environment.

In most cases of industrial production, both of these reasons – the one based on mass conservation and the one based on entropy generation – hold at the same time. Therefore, the joint product is typically a high entropy material. Due to its high entropy it will most often be considered useless and, therefore, an undesired waste; however, one should be careful to note that the classification of an output as 'waste' carries a certain value judgment, which cannot be inferred from thermodynamics alone.[5]

5.3 Thermodynamic (In)Efficiency of Industrial Production

The thermodynamic analysis in the previous section has demonstrated that the existence of a high entropy joint product is necessary and unavoidable in every process of regular industrial production. In reality, however, much of the waste currently generated is obviously avoidable. Yet this observation is not in contradiction to the result derived above. While the reference model was based on the assumption of thermodynamic efficiency, current technology and production practices are to a large extent thermodynamically inefficient. As a consequence, while a certain amount of waste is necessary and unavoidable for thermodynamic reasons, the actual amount of waste produced with current technologies is an expression of inefficiency. Thermodynamic considerations which originated in the applied field of engineering thermodynamics, in particular the exergy concept, can tell us exactly what amount of waste is due to inefficiency and may, in principle, be reduced (e.g. Ayres 1999, Bejan et al. 1996, Brodyansky et al. 1994, Cleveland and Ruth 1997, Creyts 2000, de Swaan Arons and van der Kooi 2001, de Swaan Arons et al. 2003, Dewulf et al. 2000, Ruth 1995b, 1995c).[6]

---

[5]For a review of various attempts to construct a so-called 'entropy theory of value', and a refutation of these endeavours see Baumgärtner et al. (1996).

[6]For a more general discussion of the relevance of the exergy concept for the field of *Industrial Ecology* see Connelly and Koshland (2001).

5.3.1 Engineering Thermodynamics: The Exergy Concept

Exergy is defined to be the maximum amount of work obtainable from a system as it approaches thermodynamic equilibrium with its environment in a reversible way (Ayres 1998: 192, Szargut et al. 1988: 7). Exergy is also commonly called 'available energy', or 'available work', and corresponds to the useful part of energy, thus combining the insights from both the First and Second Laws of Thermodynamics. Hence, exergy is what most people mean when they use the term 'energy', for example, when saying that 'energy is used' to carry out a certain process. As the system might consist simply of a bulk of matter, exergy is also a measure of the potential work embodied in a material, whether it is a fuel, food or other substance (Ayres et al. 1998). The exergy content of different materials can be calculated for standard values specifying the natural environment, by considering how that material eventually reaches thermodynamic equilibrium with its environment with respect to temperature, pressure, chemical potential and all other intensive variables.

The relationship between the concepts of entropy and exergy is simple, as $B_{\text{lost}} = T_0 S_{\text{gen}}$ (Law of Gouy and Stodola), where B_{lost} denotes the potential work or exergy lost by the system in a transformation process, T_0 denotes the temperature of the system's environment, and S_{gen} denotes the entropy generatedin the transformation. This means, as the system's entropy increases as a consequence of irreversible transformations according to the Second Law, the system loses exergy or some of its potential to perform work. Exergy, unlike energy, is thus not a conserved quantity. While the entropy concept stresses that with every transformation of the system something useless is created, the exergy concept stresses that something useful is diminished. These developments are two aspects of the same irreversible character of transformations of energy and matter. The character of regular industrial production, as sketched in Figure 5.1 above, therefore has a corresponding description in terms of exergy (Figure 5.2).

With the strict correspondence, established by the Law of Gouy and Stodola, between the entropy generated in an irreversible transformation and exergy lost in this process, our entire analysis could, in principle, be based either on the entropy concept or on the exergy concept. Physicists usually prefer the entropy route, as entropy is the concept traditionally established in physics. On the other hand, exergy seems to be more popular with engineers and people interested in applied work. Instead of preferring one route to the other, or going all the way along both routes in parallel, we illustrate the fruitfulness of both approaches by employing them at different stages of the argument. While we have demonstrated above the result (Proposition 5.1) that regular industrial production necessarily yields waste by-products based on the entropy concept, we now switch to the exergy concept to analyze the efficiency of regular industrial production.

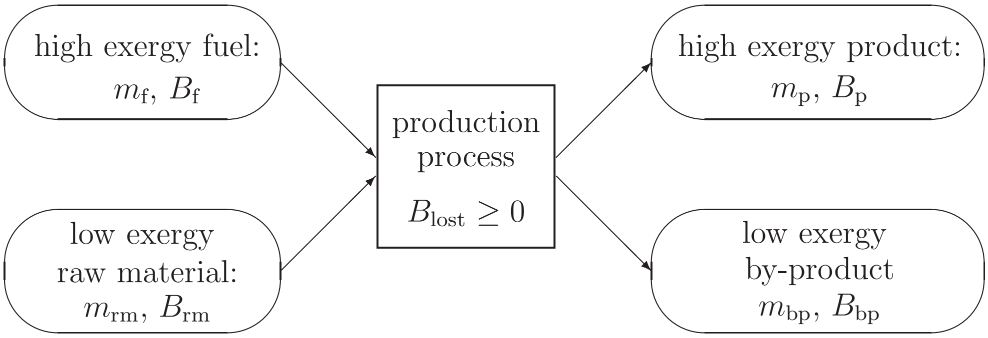

Figure 5.2 The thermodynamic structure of regular industrial production in terms of mass and exergy.

5.3.2 (In)Efficiency in Thermodynamic Equilibrium

In this section, regular industrial production is quantitatively analyzed in exergy terms with regard to thermodynamic (in)efficiency. For that sake, we turn in detail to one particular step in the production process introduced above as an illustrative example. In the production of pure iron from iron ore, the first step is to extract the ore from the deposit. In the next step, the ore is separated by physical means into iron oxide and silicates. The third step, which we shall analyze in detail in this section, then consists of chemically reducing the iron oxide to pure iron. This reduction requires exergy. It is typically provided by burning coke, which, for the purpose of this analysis, can be taken to be pure carbon. So, in the terminology outlined earlier, the desired product of this transformation is pure iron (Fe), the raw material is iron oxide (Fe_2O_3) and the fuel is carbon (C). As a waste joint product in this reaction, carbon dioxide (CO_2) is generated. The chemical reaction in this production process may be written down as follows:

$$2\ Fe_2O_3\ +\ 3\ C\ \rightarrow\ 4\ Fe\ +\ 3\ CO_2\ . \tag{5.6}$$

The molecular weight, specific entropy and exergy of the chemicals involved in the reaction are given in Table 5.1.

As one sees, the desired product Fe has a much higher exergy (i.e. lower specific entropy) than the raw material Fe_2O_3. It is the relatively high exergy content (i.e. low specific entropy) of the fuel C that provides the exergy for this transformation to happen. The waste CO_2 is then characterized by low exergy content (i.e. high specific entropy).

chemical	molecular weight [g/mole]	specific entropy [J/mole K]	exergy [kJ/mole]
Fe	56	27.3	376.4
Fe_2O_3	160	87.4	16.5
C	12	5.7	410.3
CO_2	44	213.8	19.9
O_2	32	161.1	4.0

Table 5.1 Molecular weight, specific entropy of the different chemicals involved in the reduction of iron oxide to pure iron. One mole is, by definition, the amount of some material that contains as many atoms as 12 g of carbon isotope ^{12}C*. For every material, one mole contains* 6.022×10^{23} *particles. The molecular weight of a material is its mass per mole. Source: Kondepudi and Prigogine (1998: Appendix), Szargut et al. (1988: Appendix, Table I).*

Mass balance

The chemical reaction equation (5.6) is correct in terms of the mass balance: all atoms of an element that go into the reaction come out of the reaction as well. Conservation of mass is the reason for the existence of the joint product CO_2. Producing four moles of Fe, thus, entails three moles of CO_2 as waste. That makes 0.75 moles of waste CO_2 emissions per mole of Fe produced (corresponding to 0.59 kg CO_2 per kg of Fe) for mass balance reasons alone.

Thermodynamically efficient energy balance

Checking the reaction equation (5.6) with the exergy values given in Table 5.1 reveals that while the reaction equation is written down correctly in terms of the mass balance, it is not yet correct in energetic terms. For, in order to produce four moles of Fe with an exergy content of 1,505.6 kJ, one needs the input of at least 1505.6 kJ as well. (Recall that exergy cannot be created, but always diminishes in the course of a transformation due to irreversibility.) But three moles of C only contain 1,230.9 kJ. Therefore, one actually needs more than three moles of C to deliver enough exergy for four moles of Fe to be produced from Fe_2O_3. We compensate for this shortage of exergy on the input side by introducing 0.76 additional units of the exergy source C. The reaction equation should, thus, be written down as follows to obey both laws of thermodynamics:

$$\begin{array}{ccccccccc} 2\,Fe_2O_3 & + & 3.76\,C & + & 0.76\,O_2 & \rightarrow & 4\,Fe & + & 3.76\,CO_2 \\ (33.0) & & (1,542.7) & & (3.0) & & (1,505.6) & & (74.8) \end{array} \quad . \qquad (5.7)$$

The numbers in brackets below each input and output give the exergy content in kJ of the respective amounts of inputs and outputs. On the input side 0.76

moles of oxygen (O_2) have been added to fulfill the mass balance with the additional 0.76 moles of C involved. This oxygen comes from the air and enters the transformation process when carbon is burned.

From reaction equation (5.7) we see that the exergy supplied to the reaction by its inputs (1,580 kJ) now suffices to yield the exergy of the outputs (1,580 kJ). As no exergy is lost in the reaction, that is, the exergy of the inputs exactly equals the exergy of the outputs, this corresponds to a thermodynamically 100 %-efficient and reversible transformation, in which no entropy is generated and no exergy is lost. In mass terms, reaction equation (5.7) tells us that even in thermodynamically ideal transformations of Fe_2O_3 into Fe, 3.76 moles of CO_2 are generated as material waste when producing four moles of Fe. That makes 0.94 moles of waste CO_2 emissions per mole of Fe produced (corresponding to 0.74 kg CO_2 per kg of Fe). This amount is the minimum waste generation required by the two laws of thermodynamics, as shown in Section 5.2 above to necessarily exist (Proposition 5.1).

Thermodynamic inefficiency

In real production processes the exergy content of carbon of 410.3 kJ/mole is never put to work with an efficiency of 100 %. Detailed data on pig iron production in real blast furnaces in Poland (Szargut et al. 1988: Table 7.3), where coke is burned together with atmospheric oxygen, imply that the efficiency of exergy conversion is only about 33 %.[7] This means that out of one mole of C one obtains only 135.4 kJ instead of the ideal value of 410.3 kJ. As a consequence, in order to deliver the exergy necessary to carry out the chemical reaction one needs to employ at least 12.42 moles of C. The reaction equation for a transformation that is only 33 %-efficient in energy conversion would thus read:

$$\begin{array}{ccccccccccc} 2\,\mathrm{Fe_2O_3} & + & 12.42\,\mathrm{C} & + & 9.42\,\mathrm{O_2} & \rightarrow & 4\,\mathrm{Fe} & + & 12.42\,\mathrm{CO_2} & + & \text{heat}\,. \\ (33.0) & & (5,095.9) & & (37.7) & & (1,505.6) & & (247.2) & & \end{array} \tag{5.8}$$

Out of the 5,095.9 kJ of exergy supplied by 12.42 moles of C only 33 %, or 1,681.6 kJ, are put to work in the reaction due to the inefficiency in energy conversion. The amount of exergy supplied by the inputs but not contained in the outputs of the reaction corresponds to exergy lost in the process, $B_{\text{lost}} =$ 3,413.8 kJ, which is mainly emitted from the reaction as waste heat.

From Equation (5.8) we see that due to the inefficiency in energy conversion the amount of material fuel that is necessary to drive the transformation has

[7]Typical exergy conversion efficiencies in the process industry range from values as low as 4 %, 6 % and 9 % in the production of nitric acid, oxygen and copper respectively up to values of 58 % and 63 % in the production of hydrogen and methanol (Hinderink et al. 1999: Table 1).

more than tripled. In order to produce four moles of Fe with a 33 %-efficiency one needs 12.42 moles of C (Equation 5.8) instead of just 3.76 moles in the efficient case (Equation 5.7). As a consequence, the reaction generates 12.42 moles of CO_2 (Equation 5.8) instead of just 3.76 moles in the efficient case (Equation 5.8). That makes 3.11 moles of waste CO_2 emissions per mole of Fe produced (corresponding to 2.45 kg CO_2 per kg of Fe), with only 30 % (0.94 moles) due to thermodynamic necessity (cf. the discussion of Equation 5.7 above) and 70 % (2.17 moles) due to thermodynamic inefficiency.

From this analysis one might conclude that roughly two thirds of the waste currently generated in iron production is due to thermodynamic inefficiency, while one third is actually necessary for thermodynamic reasons. Therefore, even increasing thermodynamic process efficiency to the ideal value of 100 % will not reduce the amount of waste to zero, but only to one third of the amount currently generated.

5.3.3 Finite-Time/Finite-Size Thermodynamics

Pointing to the thermodynamic inefficiency of a real production process, and how it implies the occurrence of large amounts of waste, seems to suggest that the amount of waste can easily be reduced by increasing the thermodynamic efficiency at which the process is carried out. However, there are good economic reasons why this form of thermodynamic inefficiency may actually be desired.

The analysis so far was entirely based on concepts and methods from ideal equilibrium thermodynamics, which means that a level of 100 %-efficiency in this framework is reached by operating processes in a completely reversible way between one equilibrium state and another equilibrium state, resulting in zero entropy generation (or: exergy loss) during the process. Recent research in the applied field of engineering thermodynamics has addressed the circumstance that chemical and physical processes in industry never happen in a completely reversible way between one equilibrium state and another equilibrium state. Rather, these processes are enforced by the operator of the process and they are constrained in space and time. This has led to an extension of ideal equilibrium thermodynamics, known as *finite-time/finite-size thermodynamics* (e.g. Andresen et al. 1984; Bejan 1996, 1997, Bejan et al. 1996).

From the point of view of finite-time/finite-size thermodynamics it becomes obvious that the minimum exergy requirement and minimum waste production in chemical or physical processes is considerably higher than that suggested by the ideal equilibrium thermodynamics analysis carried out so far. The reason for the increased exergy requirement (which entails an increased amount of waste at the end of the process) lies in the fact that chemical and physical transformations are forced to happen over a finite time by the operator of the production plant, which necessarily causes some dissipation of energy. In the language of the reference model described earlier, this shows in a strictly positive amount S_{gen} of entropy generated in the process.

The finite-time/finite-size consideration is a very relevant consideration for many production processes, in particular in the chemical industry. Finite-time/finite-size thermodynamics allows one to exactly identify, track down and quantify exergetic inefficiencies at the individual steps of a production processes (Bejan 1996, 1997, Bejan et al. 1996, Brodyansky et al. 1994, Creyts 2000, Szargut et al. 1988), along the entire chain of a production process (Ayres et al. 1998, Cornelissen and Hirs 1999, Cornelissen et al. 2000), and for whole industries (Dewulf et al. 2000, Hinderink et al. 1999). Thus, it yields valuable insights into the origins of exergy losses and forms a tool for designing industrial production systems in an efficient and sustainable manner (Connelly and Koshland 2001, de Swaan Arons and van der Kooi 2001, de Swaan Arons et al. 2003).

An example which demonstrates how large S_{gen} can actually be is the enrichment of uranium (Balian 1991: 347–348, 383–385). In the production of enriched uranium the actual exergy input is larger than the theoretical minimum calculated from ideal equilibrium thermodynamics by a factor of 70 million! At the Eurodif factory, the French enriching plant from which the data are taken, the process of enriching by isotope separation is realized by gas diffusion through a semipermeable membrane. An ideal process realization would require letting the gas diffuse in thermodynamic equilibrium, which would take an infinite time span. In order to carry out the process in finite time, diffusion is enhanced by building up an enormous pressure difference between the two sides of the membrane, which requires an equally enormous amount of energy. Then, the process of diffusion is no longer an equilibrium process. Instead, it is irreversible and $S_{\text{gen}} > 0$. A comparison of the ideal separation process and the real process realization shows that the huge irreversible loss of energy in the actual separation process is entirely due to the dissipation of energy in the many compressions and decompressions which are necessary to run the separation process under a pressure difference and, thus, in finite time. This dissipated energy leaves the process as waste heat.[8]

5.4 Conclusion

Our basic result is twofold. First, based on a thermodynamic analysis we have confirmed previous assertions that waste is an unavoidable and necessary joint

[8]Note that the efficiency of uranium enrichment in a centrifuge plant is about two orders of magnitude higher than in a membrane plant. Despite the tremendous exergetic inefficiency, enrichment of uranium makes sense, as the exergy loss in enrichment is small with respect to the exergy content of enriched uranium as a fuel for nuclear fission.

output in the regular industrial production of desired goods. This is the type of production technology that is currently in use and dominates production in industrial economies. Second, thermodynamic analysis has also allowed us to quantify the amount of waste that – beyond the thermodynamic minimum required – is due to inefficiencies. We have identified three major reasons for the occurrence of large amounts of excessive material waste from regular industrial production:

1. The first reason is simply *conservation of mass*. Starting with a raw material, which is a mixture of different chemical elements, to produce a desired product, which is made up of only one particular chemical element, necessarily leaves a material waste.

2. The second reason is the use of a *material fuel*, which is a characteristic property of many industrial production technologies currently in use. The fuel – carbon in our example – only serves to provide the exergy for the chemical reaction. The carbon material itself is actually neither wanted nor needed in the reaction. Because mass is conserved, the fuel material has to go somewhere after its exergy content has been stripped off. And that makes the waste. An alternative, immaterial way of providing exergy to production processes would be the use of renewable energy sources, such as solar, wind, tidal or hydro-energy.[9]

3. The third reason is the *thermodynamically inefficient performance* of current technologies when it comes to the conversion of exergy, which is a necessary factor of production in all production processes. In particular, this is due to the operation of production processes under *non-equilibrium conditions*, in order to have them completed in finite time. The shorter the time span within which one wants the process to be completed, the more energy will irreversibly be dissipated. This inefficiency not only considerably increases the need for fuel beyond the minimum exergy requirement; it also increases the amount of material waste generated far beyond the thermodynamic necessity. This holds, in particular, for carbon dioxide emissions when carbon (e.g. coal or coke) or hydrocarbons (e.g. oil or natural gas) are used as a fuel.

---

[9]Note that, because the primary goal of carrying out the transformation studied using reaction equations (5.6)–(5.8) is to split Fe_2O_3 into Fe and O_2, the minimal way of doing that would be: $2\ Fe_2O_3 + \text{direct exergy} \rightarrow 4\ Fe + 3\ O_2$. The exergy necessary to achieve the splitting of Fe_2O_3 into Fe and O_2 could, for instance, be delivered by solar energy directly. Without any material fuel the amount of material waste would be considerably reduced. With four moles of Fe there would be three moles of O_2 jointly produced. That makes 0.75 moles of waste O_2 emissions per mole of Fe produced (corresponding to 0.43 kg O_2 per kg of Fe). However, running the chemical process in this direct way, that is, powered by solar energy instead of material fuel input, would require technologies very different from the ones we are currently using.

Summing up, notwithstanding the fundamental insight that waste is a necessary joint output in regular industrial production technologies as they are currently in use, a large potential exists for the reduction of waste. Thermodynamics has proven to be a very useful analytical tool for studying and exploiting this potential. In order to translate this potential into real solutions, new production technologies are needed. In particular, new technologies should have higher exergetic fuel efficiency than existing ones. Or, even better, they should not use a material fuel at all for their exergy input, but use renewable energy sources.

Appendix

A5.1 Proof of Proposition 5.1

In order to prove the proposition (see Baumgärtner 2000: 72–77), one should distinguish between different chemical elements, such as e.g. oxygen (O), carbon (C) or iron (Fe). Each input and each output of a production process may, in general, be composed of various such elements. For example, a raw material input into production may be iron oxide (Fe_2O_3), which consists of the two chemical elements of iron (Fe) and oxygen (O). In any chemical reaction, the mass of each element is conserved separately,[10] so that the thermodynamic Law of Mass Conservation (Equation 5.4) should be formulated more precisely as

$$\begin{aligned} \forall_e \quad m_{\mathrm{rm}}(e) + m_{\mathrm{f}}(e) \quad &= \quad m_{\mathrm{p}}(e) + m_{\mathrm{bp}}(e) \\ &\quad\;\; \text{with } e = \ldots, \mathrm{O}, \ldots, \mathrm{C}, \ldots, \mathrm{Fe}, \ldots \;, \end{aligned} \tag{A5.1}$$

where $m(e)$ denotes the mass of chemical element e. For instance, in the reduction of iron oxide (Fe_2O_3) into pure iron (Fe) by means of coke (C), the mass of all chemical elements – oxygen, carbon, iron and others – is conserved separately; that is, the mass of iron in the inputs must equal the mass of iron in the outputs, and similarly for the mass of oxygen, carbon, etc.

In the proof, a mass balance will explicitly be considered for only one chemical element, the so-called ‘element under consideration’. Beyond that, it is only important that there exists more than one chemical element. But these other elements’ mass balance is of no explicit interest. In order to simplify the presentation, we therefore omit the argument e where it is clear that we refer to the first element. The detailed mass balance (A5.1) then reduces to Equation (5.4), but nevertheless is meant to refer to the first element.

[10] In principle, chemical elements may be transformed into each other by nuclear reactions. However, this possibility is neglected here.

In this interpretation, Properties (5.1) and (5.3) may be relaxed to

$$m_{\mathrm{rm}}, m_{\mathrm{p}} \geq 0, \tag{5.1$'$}$$

$$m_{\mathrm{f}} \geq 0, \tag{5.3$'$}$$

with each of these quantities, m_{rm}, m_{p} and m_{f}, being strictly positive for at least one element. That is to say, the raw material input (as well as the fuel and the desired output) does not need to contain positive amounts of *all* elements; it is only assumed to contain *at least one* element. For instance, the raw material iron oxide (Fe_2O_3) contains iron and oxygen, but no carbon. If the element under consideration in the analysis should be, say, iron, then $m_{\mathrm{rm}} > 0$; but if the element under consideration should be carbon, then $m_{\mathrm{rm}} = 0$.

The proof is now carried out by showing that either the mass m_{bp} or the entropy S_{bp} (or both) of the second output is strictly positive. Note that not necessarily both the mass *and* the entropy have to be strictly positive for an output to exist, since the mass of an output – in our formalization – will be zero if it does not contain the element under consideration. For carrying out the proof, we distinguish between the two cases that (i) the statement of Proposition 5.1 follows already from mass conservation alone and (ii) the Second Law is essential for the existence of a second output.

Joint production as a consequence of mass conservation

In many instances, the aspect of entropy is not necessary to understand why there exists a second output besides the main product. So, to start with, let us focus on the mass aspect of inputs and outputs and neglect their entropic character. Consider first the extreme case that $m_{\mathrm{p}} = 0$; the complementary case of $m_{\mathrm{p}} > 0$ will be dealt with later on. Most obviously, if $m_{\mathrm{rm}}, m_{\mathrm{f}} \geq 0$ with at least one, m_{rm} or m_{f}, strictly positive (according to Properties 5.1$'$ and 5.3$'$ and with a suitable choice of the element under consideration) and $m_{\mathrm{p}} = 0$, it follows from the mass balance (Equation 5.4) that $m_{\mathrm{bp}} = m_{\mathrm{f}} + m_{\mathrm{rm}} > 0$. This means, a joint product cogently exists.[11]

[11] Again, the assumption of $m_{\mathrm{p}} = 0$ should *not* be interpreted as production resulting in no desired product at all, or only in an immaterial one. Nor should the assumption of $m_{\mathrm{rm}} = 0$ or $m_{\mathrm{f}} = 0$ be interpreted as saying that these inputs are absent from the production process. Recall that the mass balance (Equation 5.4) refers to one particular element, say carbon, but that there are other elements as well, say iron. Then, $m_{\mathrm{p}} = 0$ only means that there is nothing of the element under consideration contained in the desired output. However, this output may well contain other elements, e.g. iron. In the example of iron production mentioned above, let carbon (C) be the element for which the mass balance is considered. Equation (5.4) can then be read as the mass balance of carbon, that is the mass of carbon in the inputs has to equal the mass of carbon in the outputs. Of course, there are also mass balances for the other elements, such as e.g. iron or oxygen, but they are of no interest for the argument. Then, $m_{\mathrm{f}} = m_{\mathrm{f}}(\mathrm{C}) > 0$ (the fuel, coke plus oxygen, contains carbon), $m_{\mathrm{rm}} = m_{\mathrm{rm}}(\mathrm{C}) = 0$ (there is

In general, mass balance considerations make the existence of at least one joint output necessary as soon as either the raw material or the fuel (or both) contain an element which is not contained in the desired product. Note that if such an argument holds for any one material of the, in general, many materials involved in a production process, then this already suffices to establish the result.

Joint production as a consequence of the Second Law

It remains to be shown that a second output necessarily exists if $m_{\mathrm{p}} > 0$. In this case, the Law of Mass Conservation alone may not suffice to prove the existence of a joint output. For instance, for $m_{\mathrm{p}} = m_{\mathrm{f}} + m_{\mathrm{rm}}$ and $m_{\mathrm{bp}} = 0$ the mass balance is fulfilled and the existence of a joint output is not immediately obvious. In this case, however, the Second Law becomes crucial in establishing the result.

In order to build the argument in this case, let us come back to the entropic character of inputs and outputs in the reference model of industrial production. As far as the mass aspect is concerned, we make the assumption that $m_{\mathrm{f}} = 0$.[12] The mass balance (Equation 5.4) then becomes

$$m_{\mathrm{rm}} = m_{\mathrm{p}} + m_{\mathrm{bp}} \ , \tag{5.2}$$

from which it follows that[13]

$$m_{\mathrm{rm}} \geq m_{\mathrm{p}} \ . \tag{5.3}$$

Consider now the entropy balance (Equation 5.5). It can be rearranged into

$$S_{\mathrm{bp}} = S_{\mathrm{f}} + \Delta S + (S_{\mathrm{rm}} - S_{\mathrm{p}}) \ . \tag{5.4}$$

The sign of the term $S_{\mathrm{rm}} - S_{\mathrm{p}}$ can be determined from considering the fraction $S_{\mathrm{rm}}/S_{\mathrm{p}}$ which is, according to the definition of specific entropy as $\sigma = S/m$, given by:

$$\frac{S_{\mathrm{rm}}}{S_{\mathrm{p}}} = \underbrace{\frac{m_{\mathrm{rm}}}{m_{\mathrm{p}}}}_{\geq 1} \cdot \underbrace{\frac{\sigma_{\mathrm{rm}}}{\sigma_{\mathrm{p}}}}_{>1} > 1 \ , \tag{5.5}$$

no carbon in the raw material, iron ore), $m_{\mathrm{p}} = m_{\mathrm{p}}(\mathrm{C}) = 0$ (the desired main product, pure iron, does not contain any carbon), and, hence, $m_{\mathrm{bp}} = m_{\mathrm{bp}}(\mathrm{C}) > 0$. In sum, the existence of the joint output, carbon dioxide, is necessary because the carbon atoms which are originally contained in the fuel are not contained in the desired product and cannot disappear either.

[12] Again, this should not be seen as restricting the generality of the treatment, since the mass balance (Equation 5.4) refers to one particular element. So, setting $m_{\mathrm{f}} = 0$ simply amounts to a suitable choice of the element under consideration. As an illustration, consider again the example of iron making and take iron (Fe) to be the element under consideration. Then $m_{\mathrm{f}} = m_{\mathrm{f}}(\mathrm{Fe}) = 0$ only means that the fuel does not contain any iron.

[13] Note that if m_{rm} should be strictly larger than m_{p}, the existence of a second output with $m_{\mathrm{bp}} > 0$ would follow immediately from mass balance considerations. Therefore, the interesting case which genuinely requires an entropy balance argument is actually $m_{\mathrm{rm}} = m_{\mathrm{p}}$.

since $m_{\mathrm{rm}} \geq m_{\mathrm{p}}$ (Equation 5.3) and $\sigma_{\mathrm{rm}} > \sigma_{\mathrm{p}}$ (Property 5.2). Hence,

$$S_{\mathrm{rm}} - S_{\mathrm{p}} > 0 \ . \tag{5.6}$$

From this and Equation (5.4) it follows that

$$S_{\mathrm{bp}} = \underbrace{S_{\mathrm{f}}}_{\geq 0} + \underbrace{\Delta S}_{\geq 0} + \underbrace{(S_{\mathrm{rm}} - S_{\mathrm{p}})}_{>0} > 0 \ .$$

This means that the existence of a joint product with entropy S_{bp} is cogently required in order to fulfill the entropy balance. It serves to take up the excess high entropy, which cannot be contained in the desired low specific entropy product.

6. Optimal Dynamic Scale and Structure of a Multi-Pollution Economy

with Frank Jöst and Ralph Winkler

6.1 Introduction

The natural environment is being damaged by the stocks of various pollutants, which are produced in different sectors of the economy, accumulate according to different dynamic relationships, and damage different environmental goods. As an example, think of the two economic sectors 'agriculture' and 'industry'. Nitrate and pesticide run-off from agricultural cultivation accumulates in groundwater and decreases its quality as drinking water (UNEP 2002); carbon dioxide emissions from fossil fuel combustion in the industrial sector accumulate in the atmosphere and contribute to global climate change (IPCC 2001). In general, the different pollutants differ in their internal dynamics, i.e. natural degradation processes, and in their harmfulness. This has implications for the optimal dynamics of both the scale and structure of the economy. By *scale* we mean the overall level of economic activity, measured by total factor input; by *structure* we mean the composition of economic activity, measured by relative factor inputs to different sectors.

In this chapter, we look into these coupled environmental-economic dynamics from a macroeconomic point of view. In particular, we are interested in the following questions: How should the macroeconomic scale and structure change over time in response to the dynamics of environmental pollution? Is this dynamic process monotonic over time, or can a trade-off between long-run and short-run considerations (e.g. lifetime versus harmfulness of pollutants) induce a non-monotonic economic dynamics? What is the time scale of economic dynamics (i.e. change of scale and structure), and how is it influenced by the different time scales and constraints of the economic and environmental systems? These questions are relevant for the current policy discussion on the sustainable biophysical scale of the aggregate economy relative to the surrounding natural environment (e.g. Arrow et al. 1995, Daly 1992a, 1996, 1999), and how economic policy should promote structural economic change as a response to changing environmental pressures (e.g. de Bruyn 1997, Winkler 2005).

We address these questions based on a model which comprises two economic sectors, each of which produces one distinct consumption good and, at the same time, gives rise to one specific pollutant. Both pollutants accumulate to stocks which display different internal dynamics, in the sense that the respective natural deterioration rates differ, and cause welfare decreasing environmental damage independently of each other. Of course, this relatively simple model cannot offer detailed policy prescriptions. However, it is detailed enough to clarify the underlying theoretical issues. In fact, we perform a *total* analysis of economy-environment interactions in a twofold manner. First, we analyze a multi-sector economy, which is fully specified in terms of resource endowment, technology, preferences and environmental quality. Second, we consider a 'disaggregate' natural environment. This goes beyond many contributions to environmental economics, where either only one (aggregate) pollutant is considered or different pollutants give rise to the same environmental problem.

Many studies in the extant literature assume that it is the *flow* of emissions which causes environmental problems. This neglects stock accumulation and, thus, an essential dynamic environmental constraint on economic action. *Stock pollution* has been taken into account by some authors (e.g. Falk and Mendelsohn 1993, Forster 1973, Luptacik and Schubert 1982, Van der Ploeg and Withagen 1991). This is usually done at a highly aggregated level, such that only one pollutant is taken into account. The case of *several stock pollutants* which all contribute to the same environmental problem (climate change) has been studied by Michaelis (1992, 1999). He is interested in finding cost-effective climate policy measures in the multi-pollution case for a given structure of the economy and does not explicitly consider the dynamics of the production side of the economy. Aaheim (1999) goes beyond Michaelis in that he analyzes numerically the dynamics of a two-sector economy which gives rise to three different stock pollutants and which is constrained by an exogenously given policy target concerning the aggregate level of pollution. Moslener and Requate (2001) challenge the global warming potential as a useful indicator when there are many interacting greenhouse gases with different dynamic characteristics. Faber and Proops (1998: Chapter 11) and Keeler et al. (1972) explicitly study the dynamics of different production sectors with pollution, assuming one single pollutant. Winkler (2005) analyzes optimal structural change of a two-sector economy characterized by two stock quantities: the capital stock and the stock of a pollutant which is emitted from the more capital-intense sector. Baumgärtner and Jöst (2000) study the optimal (static) structure of a vertically integrated two-sector economy where both sectors produce a specific by-product. The first sector's by-product can be used as a secondary resource in the second sector.

In this chapter, we determine the optimal dynamic scale and structure of a multi-pollution economy within an optimal control framework. We use a linear approximation around the steady-state to obtain analytical results, and a numerical optimization of the non-approximated system to check for their robustness. The methodological innovation of our analysis is that we derive a closed form solution to the intertemporal optimization problem, which includes explicit expressions for the time scale of economic dynamics and the point in time where a non-monotonicity may occur. Our analysis shows that along the optimal time-path (i) the overall scale of economic activity may be less than maximal; (ii) the time scale of economic dynamics is mainly determined by the lifetime of pollutants, their harmfulness and the discount rate; and (iii) the control of economic scale and structure may be non-monotonic.

Although our modeling approach is inspired by Ramsey-type optimal growth models, which have previously been used to study steady state growth with environmental pollution (e.g. Gradus and Smulders 1993, 1996, Jöst et al. 2004, Keeler et al. 1972, Plourde 1972, Siebert 2004, Smith 1977, Van der Ploeg and Withagen 1991), we are essentially concerned with the issue of dynamic change in both scale and structure of economic activity. Therefore, in this chapter we do not restrict the analysis to steady states but focus on the explicit time-dependence of the solution. Furthermore, we study an economy without any potential for steady state growth, as this highlights the structural-change-effect, which may be obscured by growth effects otherwise. The sole genuine generator of dynamics in our model is the accumulation of pollutant stocks in the natural environment.

The chapter is organized as follows. In Section 6.2 we present the model. Section 6.3 is devoted to a formal analysis of the optimal dynamic scale and structure of the economy, based on a linear approximation around the stationary state. Section 6.4 confirms the analytical results thus obtained by a numerical optimization of the non-approximated system. Section 6.5 concludes.

6.2 The Model

We study a two sector economy with one scarce non-accumulating factor of production, say labor, two consumption goods, and two pollutants that accumulate to stocks. Welfare is determined by the amounts consumed of both consumption goods, as well as by the environmental damage caused by the two pollutant stocks.

The production of consumption goods in sectors 1 and 2 of the economy is described by two production functions, $y_i = P^i(l_i)$ for $i = 1, 2$, where l_i denotes the amount of labor allocated in sector i. With index l denoting derivatives with respect to the sole argument l_i, $P^i_l \equiv dP^i/dl_i$ and $P^i_{ll} \equiv d^2P^i/dl_i^2$, the production functions are assumed to exhibit the following standard properties:

$$P^i(0) = 0 \ , \quad P_l^i > 0 \ , \quad \lim_{l_i \to 0} P_l^i = +\infty \ , \quad P_{ll}^i < 0 \qquad (i = 1, 2) \ . \tag{6.1}$$

Since we want to analyze an economy without potential for steady state growth, we assume a fixed supply of labor, $\lambda > 0$. Consumption possibilities are described by

$$y_i = P^i(l_i) \qquad (i = 1, 2) \ , \tag{6.2}$$

$$l_1 + l_2 \leq \lambda \ . \tag{6.3}$$

In addition to the consumption good, each sector yields a pollutant which comes as a joint output in a fixed proportion to the desired output. Without loss of generality,

$$e_i = y_i \qquad (i = 1, 2) \ . \tag{6.4}$$

Both flows of pollutants, e_1 and e_2, add to the respective stock of the pollutant, which deteriorates at the constant rate δ_i:[1]

$$\dot{s}_i = e_i - \delta_i s_i \qquad \text{with } \delta_i > 0 \qquad (i = 1, 2) \ . \tag{6.5}$$

Instantaneous social welfare V depends on consumption of both goods, y_1 and y_2, and on the damage to environmental quality which hinges upon the stocks of pollutants s_1 and s_2. We consider the following welfare function:

$$V(y_1, y_2, s_1, s_2) = U(y_1, y_2) - \left[\frac{\sigma_1}{2} s_1^2 + \frac{\sigma_2}{2} s_2^2\right] \qquad \text{with } \sigma_1, \sigma_2 > 0 \ , \tag{6.6}$$

where σ_i indicates the harmfulness of pollutant i $(i = 1, 2)$ and U represents welfare gains due to consumption. The function U is assumed to exhibit the usual property of positive and decreasing marginal welfare in both consumption goods. In order to have an additively separable welfare function in all four arguments (y_1, y_2, s_1, s_2), we assume that neither consumption good influences marginal welfare of the other. With index i denoting the partial derivative with respect to argument y_i, i.e. $U_i \equiv \partial U / \partial y_i$ and $U_{ij} \equiv \partial^2 U / \partial y_i \partial y_j$ with $i, j = 1, 2$, the assumptions are:

$$U_i > 0 \ , \quad \lim_{y_i \to 0} U_i = +\infty \ , \quad U_{ii} < 0 \ , \quad U_{ij} = 0 \qquad (i, j = 1, 2 \text{ and } i \neq j) \ . \tag{6.7}$$

Both stocks of pollutants exert an increasing marginal damage, which is captured in the welfare function V, for the sake of tractability, by quadratic damage functions. Furthermore, both stocks decrease welfare independently. This is

[1] In general, the decay rate may depend on emissions and the stock: $\delta_i = \delta_i(e_i, s_i)$. For analytical tractability, we assume δ_i to be constant.

plausible if they damage different environmental goods. Thus, the welfare effect of one additional unit of one pollutant does not depend on the amount of the other. Note that the overall welfare function V is strictly concave.

Since we are interested in studying questions related to the scale as well as the structure of economic activity, and in order to simplify the analysis of corner solutions in the optimization problem, we introduce new dimensionless variables in the following way:

$$c = \frac{l_1 + l_2}{\lambda} \qquad \text{and} \qquad x = \frac{l_1}{l_1 + l_2} \ . \tag{6.8}$$

The variable c stands for the *scale* of economic activity. It indicates what fraction of the total available amount of labor is devoted to economic activity, and may take values between 0 and 1. The remaining fraction $1 - c$ is left idle. This can be interpreted as an implicit form of pollution abatement. By not using all available labor in the production of the consumption goods (and, consequently, emissions) but leaving part of the labor endowment idle, the variable c can be thought of as measuring the scale of economic activity *in the sectors producing consumption goods and pollution*, whereas the fraction $1-c$ of labor may be thought of as being employed in (implicit) pollution abatement.[2] The variable x stands for the *structure* of economic activity. It indicates the fraction of the total labor employed in production, $l_1 + l_2$, that is allocated to sector 1, and may take values between 0 and 1. The remaining fraction $1 - x$ is allocated to sector 2. The variables l_1 and l_2 can then be expressed in terms of c and x:

$$l_1 = l_1(c, x) = cx\lambda \qquad \text{and} \qquad l_2 = l_2(c, x) = c(1 - x)\lambda \ .$$

This allows us to replace l_1 and l_2 in the problem. For notational convenience, we introduce new production functions F^i which depend directly on c and x, and which are defined in the following way:

$$F^i(c, x) \equiv P^i(l_i(c, x)) \qquad \text{for all } c,\ x\ . \tag{6.9}$$

From (6.1) and (6.9) one obtains that the F^i have the following properties:

$$F^1_c = xP^1_l\lambda > 0\ , \qquad \lim_{c\to 0} F^1_c(x \neq 0) = +\infty\ , \tag{6.10}$$

$$F^1_x = cP^1_l\lambda > 0\ , \tag{6.11}$$

$$F^2_c = (1 - x)P^2_l\lambda > 0\ , \qquad \lim_{c\to 0} F^2_c(x \neq 1) = +\infty\ , \tag{6.12}$$

$$F^2_x = -cP^2_l\lambda < 0\ . \tag{6.13}$$

---

[2]Not taking into account potential abatement activities for the scale of economic activity is in line with arguments from the 'green national product' discussion, according to which defensive and restorative activities should not be counted as augmenting the net national product (e.g. Ahmad et al. 1989, World Bank 1997).

6.3 Optimal Scale and Structure of the Economy

Taking a social planner's perspective, we now determine the optimal scale and structure of the multi-pollution economy described in the previous section. The control variables are the scale (c) and the structure (x) of economic activity. In terms of pollution, the choice over c and x is a choice over (i) how much pollution to emit overall, and (ii) what particular pollutant to emit. These are the two essential macroeconomic dimensions of every multi-pollution allocation decision.

6.3.1 Intertemporal Optimization

We maximize the discounted intertemporal welfare over c and x,

$$\int_0^\infty \left[U(y_1, y_2) - \frac{\sigma_1}{2} s_1^2 - \frac{\sigma_2}{2} s_2^2 \right] e^{-\rho t} dt \ , \tag{6.14}$$

where ρ denotes the discount rate and $y_i = F^i(c, x)$ $(i, j = 1, 2)$, subject to the dynamic constraints for the two state variables s_1 and s_2 which are given by Equations (6.5):

$$\dot{s}_i = F^i(c, x) - \delta_i s_i \qquad \text{with} \ \ \delta_i > 0 \qquad (i = 1, 2) \ . \tag{6.15}$$

In addition, the following restrictions for the control variables c and x hold:

$$0 \leq c \leq 1 \qquad \text{and} \qquad 0 \leq x \leq 1 \ . \tag{6.16}$$

Corner solutions with $x = 0$ or $x = 1$ cannot be optimal since either case would imply, due to Assumptions (6.1) and (6.7), that the marginal utility of one consumption good would go to infinity while the marginal utility of the other would remain finite. Similarly, a corner solution with $c = 0$ cannot be optimal since in that case the marginal utility of both consumption goods would go to infinity while the marginal damage from environmental pollution would remain finite. Hence, the only remaining restriction, which we have to control for explicitly, is:

$$c \leq 1 \ . \tag{6.17}$$

We introduce two costate variables, p_1 and p_2, and a Kuhn-Tucker parameter, p_c. The current value Hamiltonian of the problem then reads

$$\begin{aligned} \mathcal{H}(c, x, s_1, s_2; p_1, p_2, p_c) \ = \ & U(F^1(c, x), F^2(c, x)) - \frac{\sigma_1}{2} s_1^2 - \frac{\sigma_2}{2} s_2^2 \\ & + p_1 \left[F^1(c, x) - \delta_1 s_1 \right] \\ & + p_2 \left[F^2(c, x) - \delta_2 s_2 \right] \\ & + p_c \left[1 - c \right] \ . \end{aligned} \tag{6.18}$$

Since both control variables, c and x, are always strictly positive, the two state variables, s_1 and s_2, are always nonnegative and the Hamiltonian $\mathcal{H}$ is continuously differentiable with respect to c and x, the first order conditions of the control problem are:

$$
\begin{aligned}
U_1 F_c^1 \;+\; U_2 F_c^2 \;+\; p_1 F_c^1 \;+\; p_2 F_c^2 \;-\; p_c \;&=\; 0 \;, && (6.19)\\
U_1 F_x^1 \;+\; U_2 F_x^2 \;+\; p_1 F_x^1 \;+\; p_2 F_x^2 \;&=\; 0 \;, && (6.20)\\
\sigma_1 s_1 \;+\; (\delta_1 + \rho) p_1 \;&=\; \dot{p}_1 \;, && (6.21)\\
\sigma_2 s_2 \;+\; (\delta_2 + \rho) p_2 \;&=\; \dot{p}_2 \;, && (6.22)\\
p_c \geq 0 \;,\; p_c(1 - c) \;&=\; 0 \;, && (6.23)
\end{aligned}
$$

plus the dynamic constraints (6.15) and the restriction (6.17). These necessary conditions are also sufficient if, in addition, the transversality conditions

$$\lim_{t \to \infty} p_i(t)\, e^{-\rho t} \cdot s_i(t) = 0 \qquad (i = 1, 2) \;, \tag{6.24}$$

hold (see Appendix A6.1). Note that the optimal path is also unique.

6.3.2 Stationary State

Setting $\dot{p}_1 = 0$, $\dot{p}_2 = 0$, $\dot{s}_1 = 0$ and $\dot{s}_2 = 0$ in the system of first order conditions (6.15), (6.17) and (6.19)–(6.23) yields the necessary and sufficient conditions for an optimal stationary state $(c^\star, x^\star, s_1^\star, s_2^\star)$, in which neither the scale nor the structure of economic activity nor the stocks of pollution accumulated in the environment change over time. From conditions (6.21) and (6.22) one obtains for the costate variables p_i $(i = 1, 2)$:

$$p_i = -\frac{\sigma_i s_i^\star}{\delta_i + \rho} \qquad (i = 1, 2) \;. \tag{6.25}$$

Inserting (6.25) in (6.19) and (6.20), and rearranging terms, yields the following necessary and sufficient conditions for an optimal stationary state:

$$
\begin{aligned}
U_1^\star - \frac{\sigma_1 s_1^\star}{\delta_1 + \rho} \;&=\; \frac{p_c F_x^{2\star}}{F_c^{1\star} F_x^{2\star} - F_x^{1\star} F_c^{2\star}} \;, && (6.26)\\
U_2^\star - \frac{\sigma_2 s_2^\star}{\delta_2 + \rho} \;&=\; \frac{-p_c F_x^{1\star}}{F_c^{1\star} F_x^{2\star} - F_x^{1\star} F_c^{2\star}} \;, && (6.27)
\end{aligned}
$$

where $U_i^\star$ and $F_j^{i\star}$ $(i = 1, 2\,;\, j = c, x)$ denote functions evaluated at stationary state values of the argument. From the signs of the F_j^i and p_c stated in (6.10)–(6.13) and (6.23), it follows that:

$$U_i^\star \geq \frac{\sigma_i s_i^\star}{\delta_i + \rho} \qquad (i = 1, 2) \;, \tag{6.28}$$

where the ">" sign indicates a corner solution ($c^\star = 1$). Furthermore, from the equations of motion (6.15) one obtains

$$s_i^\star = \frac{F^{i\star}}{\delta_i} = \text{const.} \qquad (i = 1, 2)\ . \tag{6.29}$$

The interpretation of the two conditions (6.28) is that in an interior (corner) optimal stationary state the scale and structure of economic activity are such that for each sector the marginal welfare gain due to consumption of that sector's output equals (is greater than) the aggregate future marginal damage from that sector's current emission which comes as an inevitable by-product with the consumption good.[3]

An optimal stationary state exists if the system (6.23), (6.26), (6.27) and (6.29) of five equations for the five unknowns ($c^\star, x^\star, s_1^\star, s_2^\star$) and $p_c^\star$ has a solution with $0 < c^\star \leq 1$ and $0 < x^\star < 1$. With the properties of the utility and production functions assumed here, a unique optimal stationary state always exists.

Proposition 6.1

(i) There exists a unique stationary state ($c^\star, x^\star, s_1^\star, s_2^\star$), which is given as the solution to (6.23), (6.26), (6.27) and (6.29).

(ii) The optimal stationary state of the economy is an interior solution with $c^\star < 1$, if the total available amount of labor λ in the economy is strictly greater than some threshold value $\bar{\lambda} = \bar{l}_1 + \bar{l}_2$, where the $\bar{l}_i$ are specified by the following implicit equations:

$$U_i(P^1(\bar{l}_1), P^2(\bar{l}_2)) = \frac{\sigma_i P^i(\bar{l}_i)}{\delta_i^2 + \delta_i \rho} \qquad (i = 1, 2)\ .$$

Proof: see Appendix A6.2.

In the following, we shall concentrate on the case of an interior stationary state with $c^\star < 1$. Hence, we assume that the total labor amount λ exceeds $\bar{\lambda}$ as specified in Proposition 6.1. In order to study the properties of the interior optimal stationary state ($c^\star, x^\star$) some comparative statics can be done with Conditions (6.26), (6.27) and (6.29). The results are stated in the following proposition.

---

[3]Note that taking account of discounting and the natural degradation of the respective pollution stock, the net present value of the accumulated damage of one marginal unit of pollution sums up to the right-hand-side of (6.28), as $\int_0^\infty \sigma_i s_i^\star e^{-(\rho+\delta_i)t} dt = \sigma_i s_i^\star / (\rho + \delta_i)$ $(i = 1, 2)$.

Proposition 6.2
An interior optimal stationary state, if it exists, has the following properties:

$$\begin{array}{llll}
\dfrac{dc^\star}{d\delta_1} > 0 \ , & \dfrac{dx^\star}{d\delta_1} > 0 \ , & \dfrac{dc^\star}{d\delta_2} > 0 \ , & \dfrac{dx^\star}{d\delta_2} < 0 \ , \\
\dfrac{dc^\star}{d\sigma_1} < 0 \ , & \dfrac{dx^\star}{d\sigma_1} < 0 \ , & \dfrac{dc^\star}{d\sigma_2} < 0 \ , & \dfrac{dx^\star}{d\sigma_2} > 0 \ , \\
\dfrac{dc^\star}{d\rho} > 0 \ , & \dfrac{dx^\star}{d\rho} \gtreqless 0 \quad \text{for} & \multicolumn{2}{l}{\dfrac{[U_{22}^\star \delta_2(\delta_2+\rho)-\sigma_2](\delta_2+\rho)}{[U_{11}^\star \delta_1(\delta_1+\rho)-\sigma_1](\delta_1+\rho)} \gtrless \dfrac{\sigma_2 F^{2\star} F_c^{1\star}}{\sigma_1 F^{1\star} F_c^{2\star}} \ .}
\end{array}$$

Proof: see Appendix A6.3.

These results can be interpreted as follows. For both pollutants i ($i = 1, 2$), the lower is the natural deterioration rate δ_i and the higher is the harmfulness σ_i, the lower is the relative weight of the emitting sector in the total economy and the lower is the overall scale of economic activity in the stationary state. An increase in the discount rate ρ increases the optimal stationary scale of economic activity, $c^\star$, while its effect on the optimal stationary structure of economic activity, $x^\star$, is ambiguous.

6.3.3 Optimal Dynamic Path and Local Stability Analysis

In the following we solve the optimization problem by linearizing the resulting system of differential equations around the stationary state. Since our model is characterized by only mild non-linearities,[4] we expect the linear approximation to yield insights which should also hold for the exact problem. In Section 6.4 below, we shall numerically optimize the exact problem, and confirm this expectation.

As we have assumed an interior stationary state, the optimal path will also be an interior optimal path at least in a neighborhood of the interior stationary state. Hence, we restrict the analysis to the case of an interior solution, i.e. $c^\star < 1$. As shown in Appendix A6.4, the optimal dynamics of the two control variables c, x and the two state variables s_1, s_2 can be described by a system of four coupled first order autonomous differential equations:

$$\dot{c} = \frac{[U_1(\delta_1+\rho)-\sigma_1 s_1]U_{22}F_x^2 - [U_2(\delta_2+\rho)-\sigma_2 s_2]U_{11}F_x^1}{U_{11}U_{22}df} \ , \tag{6.30}$$

$$\dot{x} = \frac{[U_2(\delta_2+\rho)-\sigma_2 s_2]U_{11}F_c^1 - [U_1(\delta_1+\rho)-\sigma_1 s_1]U_{22}F_c^2}{U_{11}U_{22}df} \ , \tag{6.31}$$

$$\dot{s}_1 = F^1 - \delta_1 s_1 \ , \tag{6.32}$$

$$\dot{s}_2 = F^2 - \delta_2 s_2 \ , \tag{6.33}$$

[4]Remember that the welfare function V is additively separable in all four arguments.

with $df \equiv F_c^1 F_x^2 - F_x^1 F_c^2 < 0$. Linearizing around the stationary state $(c^\star, x^\star, s_1^\star, s_2^\star)$ yields the following approximated dynamic system (see Appendix A6.5):

$$\begin{pmatrix} \dot{c} \\ \dot{x} \\ \dot{s_1} \\ \dot{s_2} \end{pmatrix} \approx J^\star \begin{pmatrix} c - c^\star \\ x - x^\star \\ s_1 - s_1^\star \\ s_2 - s_2^\star \end{pmatrix} \quad \text{with} \tag{6.34}$$

$$J^\star = \begin{pmatrix} \rho + \frac{\delta_1 F_c^{1\star} F_x^{2\star} - \delta_2 F_x^{1\star} F_c^{2\star}}{df^\star} & \frac{(\delta_1 - \delta_2) F_x^{1\star} F_x^{2\star}}{df^\star} & -\frac{\sigma_1 F_x^{2\star}}{U_{11}^\star df^\star} & \frac{\sigma_2 F_x^{1\star}}{U_{22}^\star df^\star} \\ \frac{(\delta_2 - \delta_1) F_c^{1\star} F_c^{2\star}}{df^\star} & \rho + \frac{\delta_2 F_c^{1\star} F_x^{2\star} - \delta_1 F_x^{1\star} F_c^{2\star}}{df^\star} & \frac{\sigma_1 F_c^{2\star}}{U_{11}^\star df^\star} & -\frac{\sigma_2 F_c^{1\star}}{U_{22}^\star df^\star} \\ F_c^{1\star} & F_x^{1\star} & -\delta_1 & 0 \\ F_c^{2\star} & F_x^{2\star} & 0 & -\delta_2 \end{pmatrix} .$$

The Jacobian evaluated at the stationary state, $J^\star$, has four real eigenvalues (see Appendix A6.5), two of which are strictly negative (ν_1, ν_2) and two of which are strictly positive (ν_3, ν_4). Hence, the system dynamics exhibits saddlepoint stability, i.e. for all initial stocks of pollutants, s_1^0 and s_2^0, there exists a unique optimal path which asymptotically converges towards the stationary state. Because of the transversality conditions (6.24) the optimal path is restricted to the stable hyperplane, which is spanned by the eigenvectors associated with the negative eigenvalues. Given the eigenvalues and the eigenvectors, which are calculated in Appendix A6.5, the explicit system dynamics in a neighborhood around the stationary state is given by:

$$\begin{aligned} c(t) &= c^\star + (s_1^0 - s_1^\star) \frac{F_x^{2\star}(\nu_1 + \delta_1)}{F_c^{1\star} F_x^{2\star} - F_x^{1\star} F_c^{2\star}} e^{\nu_1 t} - \\ &\quad (s_2^0 - s_2^\star) \frac{F_x^{1\star}(\nu_2 + \delta_2)}{F_c^{1\star} F_x^{2\star} - F_x^{1\star} F_c^{2\star}} e^{\nu_2 t} , & (6.35) \\ x(t) &= x^\star - (s_1^0 - s_1^\star) \frac{F_c^{2\star}(\nu_1 + \delta_1)}{F_c^{1\star} F_x^{2\star} - F_x^{1\star} F_c^{2\star}} e^{\nu_1 t} + \\ &\quad (s_2^0 - s_2^\star) \frac{F_c^{1\star}(\nu_2 + \delta_2)}{F_c^{1\star} F_x^{2\star} - F_x^{1\star} F_c^{2\star}} e^{\nu_2 t} , & (6.36) \\ s_1(t) &= s_1^\star + (s_1^0 - s_1^\star)\, e^{\nu_1 t} , & (6.37) \\ s_2(t) &= s_2^\star + (s_2^0 - s_2^\star)\, e^{\nu_2 t} , & (6.38) \end{aligned}$$

where $s_i^0 = s_i(0)$ $(i = 1, 2)$ denote the initial pollutant stocks.

As a measure of the overall rate of convergence of a process $z(t)$ which asymptotically approaches $z^\star$, we define the *characteristic time scale of convergence* τ_z by

$$\tau_z^{\ -1} \equiv \overline{\left| \frac{\dot{z}(t)}{z(t) - z^\star} \right|} , \tag{6.39}$$

where the horizontal bar denotes the average over time. The greater is the time scale τ_z, the slower is the convergence towards $z^\star$. With this definition, it is obvious from Equations (6.37) and (6.38) that the pollutant stock s_i $(i = 1, 2)$ converges towards its stationary state value $s_i^\star$ with a characteristic time scale $\tau_{s_i} = 1/\,|\nu_i|$. As the system approaches the stationary state for $t \to \infty$, the scale c and structure x (Equations 6.35 and 6.36) converge towards their stationary state values $c^\star$ and $x^\star$ with a characteristic time scale which is determined by the eigenvalue with the smaller absolute value, $\tau_c = \tau_x = 1/\min\{|\nu_1|, |\nu_2|\}$ (see Appendix A6.6). Proposition 6.3 summarizes these results.

Proposition 6.3
For the linear approximation (6.34) around the stationary state $(c^\star, x^\star, s_1^\star, s_2^\star)$ the following statements hold:

(i) The stationary state is saddlepoint-stable.

(ii) The explicit system dynamics is given by Equations (6.35)–(6.38).

(iii) The characteristic time scale of convergence towards the stationary state is given by

- $\tau_c = \tau_x = 1/\min\{|\nu_1|, |\nu_2|\}$ *for the control variables c and x, and by*
- $\tau_{s_i} = 1/\,|\nu_i|$ *for stock variable s_i $(i = 1, 2)$.*

As shown in Appendix A6.5 the eigenvalues ν_1 and ν_2 are given by

$$\nu_1 = \frac{1}{2}\left[\rho - \sqrt{(\rho + 2\delta_1)^2 - \frac{4\sigma_1}{U_{11}^\star}}\right] < 0\,, \tag{6.40}$$

$$\nu_2 = \frac{1}{2}\left[\rho - \sqrt{(\rho + 2\delta_2)^2 - \frac{4\sigma_2}{U_{22}^\star}}\right] < 0\,. \tag{6.41}$$

Hence, the absolute value of ν_i (time scale of convergence) decreases (increases) with the discount rate ρ and the curvature of consumption welfare in the stationary state $|U_{ii}^\star|$ $(i = 1, 2)$. It increases (decreases) with the harmfulness σ_i and the deterioration rate δ_i of the pollutant stock.

We now turn to the question of the (non-)monotonicity of the optimal path. According to Equations (6.37) and (6.38), the stocks of the two pollutants converge monotonically towards their stationary state values $s_1^\star$ and $s_2^\star$. In order to show that the optimal paths for the control variables c and x may be non-monotonic, we differentiate Equations (6.35) and (6.36) with respect to t:

$$\begin{aligned}\dot{c}(t) &= \nu_1(s_1^0 - s_1^\star)\frac{F_x^{2\star}(\nu_1+\delta_1)}{F_c^{1\star}F_x^{2\star} - F_x^{1\star}F_c^{2\star}}\, e^{\nu_1 t} - \\ &\quad \nu_2(s_2^0 - s_2^\star)\frac{F_x^{1\star}(\nu_2+\delta_2)}{F_c^{1\star}F_x^{2\star} - F_x^{1\star}F_c^{2\star}}\, e^{\nu_2 t}\,, \qquad (6.42)\\ \dot{x}(t) &= -\nu_1(s_1^0 - s_1^\star)\frac{F_c^{2\star}(\nu_1+\delta_1)}{F_c^{1\star}F_x^{2\star} - F_x^{1\star}F_c^{2\star}}\, e^{\nu_1 t} + \\ &\quad \nu_2(s_2^0 - s_2^\star)\frac{F_c^{1\star}(\nu_2+\delta_2)}{F_c^{1\star}F_x^{2\star} - F_x^{1\star}F_c^{2\star}}\, e^{\nu_2 t}\,. \qquad (6.43)\end{aligned}$$

The optimal path is non-monotonic if $\dot{c}$ or $\dot{x}$ change their sign, i.e. if the paths $c(t)$ or $x(t)$ exhibit a local extremum for positive times t. According to the signs of the ν_i and F_j^i ($i = 1, 2$ and $j = c, x$) and given that $\nu_1 \neq \nu_2$, $c(t)$ exhibits a unique local extremum if $\operatorname{sgn}(s_1^0 - s_1^\star) \neq \operatorname{sgn}(s_2^0 - s_2^\star)$, and $x(t)$ exhibits a unique local extremum if $\operatorname{sgn}(s_1^0 - s_1^\star) = \operatorname{sgn}(s_2^0 - s_2^\star)$.[5] Solving $\dot{c}(t) = 0$ and $\dot{x}(t) = 0$ for t, using expressions (6.42) and (6.43) for $\dot{c}$ and $\dot{x}$, yields:

$$\hat{t} = \begin{cases} \ln\left[\frac{\nu_2(s_2^0-s_2^\star)F_x^{1\star}(\nu_2+\delta_2)}{\nu_1(s_1^0-s_1^\star)F_x^{2\star}(\nu_1+\delta_1)}\right](\nu_1-\nu_2)^{-1}\,, & \text{if} \quad \operatorname{sgn}(s_1^0-s_1^\star) \neq \operatorname{sgn}(s_2^0-s_2^\star) \\ \ln\left[\frac{\nu_2(s_2^0-s_2^\star)F_c^{1\star}(\nu_2+\delta_2)}{\nu_1(s_1^0-s_1^\star)F_c^{2\star}(\nu_1+\delta_1)}\right](\nu_1-\nu_2)^{-1}\,, & \text{if} \quad \operatorname{sgn}(s_1^0-s_1^\star) = \operatorname{sgn}(s_2^0-s_2^\star) \end{cases}. \qquad (6.44)$$

According to this equation, it is possible that $\hat{t}$ may be negative or infinite, which is meaningless in the context of this analysis. In this case we would observe monotonic optimal paths for both control variables c and x for times $0 < t < +\infty$. For instance, $\hat{t}$ is negative if $|s_2^0 - s_2^\star|$ is sufficiently small, that is, the second pollutant stock is initially already close to its stationary state level. Furthermore, $\hat{t}$ equals (plus or minus) infinity if either $|s_1^0 - s_1^\star| = 0$ or $|\nu_1 - \nu_2| = 0$, that is, the first pollutant stock is initially already at its stationary state level or the eigenvalues are identical. The following proposition summarizes the behavior of the optimal control path.

Proposition 6.4
In the linear approximation (6.34) around the stationary state $(c^\star, x^\star, s_1^\star, s_2^\star)$, the following statements hold for the optimal path:

(i) The stocks of pollutants $s_1(t)$ and $s_2(t)$ converge exponentially, and hence monotonically, towards their stationary state values $s_1^\star$ and $s_2^\star$.

(ii) If and only if $\hat{t}$ as given by Equation (6.44) is strictly positive and finite, then the optimal control is non-monotonic over time and $\hat{t}$ denotes the

[5] Note that $\nu_i + \delta_i < 0$, which can easily be verified from Equations (A6.26) and (A6.27).

time at which the optimal control has a unique local extremum. In particular, if $sgn(s_1^0 - s_1^\star) \neq sgn(s_2^0 - s_2^\star)$, $c(t)$ is non-monotonic and $x(t)$ is monotonic. If $sgn(s_1^0 - s_1^\star) = sgn(s_2^0 - s_2^\star)$, $x(t)$ is non-monotonic and $c(t)$ is monotonic.

6.4 Numerical Optimization

In this section we illustrate the results derived in Section 6.3 by numerical optimizations of the original, non linearized optimization problem (6.14)–(6.16). The results thus obtained confirm that the insights from analyzing the linearized system also hold for the exact solution. All numerical optimizations were carried out with the advanced optimal control software package MUSCOD-II (Diehl et al. 2001), which exploits the multiple shooting state discretization (Leineweber et al. 2003).

There are four different qualitative scenarios which have to be examined. (i) Both stocks of pollutants exhibit the same harmfulness but differ in their deterioration rates, i.e. $\sigma_1 = \sigma_2$, $\delta_1 < \delta_2$. (ii) The two pollutants differ in their harmfulness but have equal deterioration rates, i.e. $\sigma_1 < \sigma_2$, $\delta_1 = \delta_2$. (iii) The pollutants differ in both harmfulness and deterioration rates and the more harmful pollutant has the higher deterioration rate, i.e. $\sigma_1 < \sigma_2$, $\delta_1 < \delta_2$. (iv) Both harmfulness and deterioration rates are different, and the more harmful pollutant has a lower deterioration rate, i.e. $\sigma_1 < \sigma_2$, $\delta_1 > \delta_2$. Furthermore, each of the four scenarios splits into four subcases, depending on the initial stocks of pollutants (both initial stocks below, only first stock above, only second stock above and both stocks above the stationary state levels).

In the following we discuss these four different scenarios. The parameter values used for the numerical optimization have been chosen so as to illustrate clearly the different effects, and do not necessarily reflect the characteristics of real environmental pollution problems. For all numerical examples, the total labor supply λ has been chosen so as to guarantee an interior stationary state scale $c^\star < 1$. As it is not possible to optimize numerically over an infinite time horizon, the time horizon has been set to 250 years and all parameters have been chosen in such a way that the system at time $t = 250$ is very close to the stationary state. For a more convenient exposition, the figures show the time paths up to $t = 125$ only. The parameter values for the numerical optimization are listed in Appendix A6.7.

In the first scenario ($\sigma_1 = \sigma_2$), both stocks of pollutants exhibit the same harmfulness but the deterioration rate is smaller for the first pollutant than for the second. Figure 6.1 shows the result of a numerical optimization of this case. In this example the initial stocks for both pollutants are above their stationary state levels ($s_1^0 = 30, s_2^0 = 30$). The optimal path for the structure exhibits non-monotonic behavior as expected from Proposition 6.4. Further, we expect that the optimal stationary state structure $x^\star$ is clearly below 0.5, indicating that relatively more labor is employed in the second sector, because as the second stock of pollutant deteriorates at a higher rate the aggregate intertemporal damage of one unit of emissions is smaller for the second pollutant.[6] This expectation is confirmed by the numerical optimization.

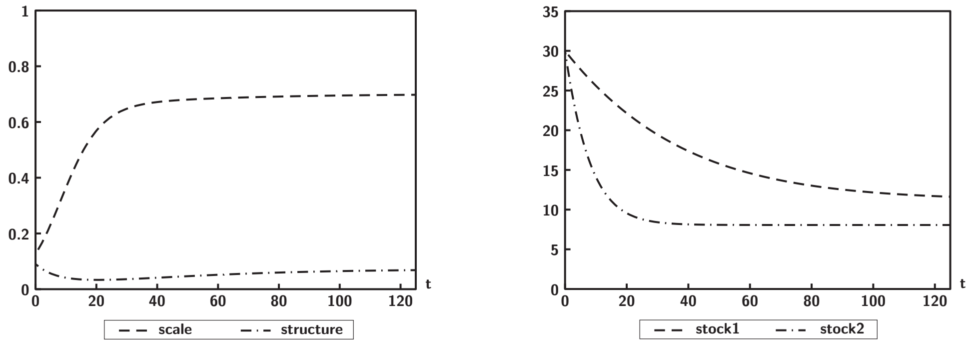

Figure 6.1 Optimal paths for scale and structure (left) and the two pollutant stocks (right) for the case $\sigma_1 = \sigma_2$, $\delta_1 < \delta_2$. Parameter values used for the numerical optimization are given in Appendix A6.7.

In the second scenario ($\sigma_1 < \sigma_2, \delta_1 = \delta_2$), the two stocks of pollutants are of different harmfulness but the deterioration rate for the two pollutants are equal. The result of a numerical optimization of this case is presented in Figure 6.2. In this example the initial stock for the first (second) pollutant is above (below) their stationary state levels ($s_1 = 40, s_2 = 0$). Now, the optimal path for the scale exhibits a non monotonic behavior as expected from Proposition 6.4. Further, we expect that the optimal stationary state structure $x^\star$ is clearly above 0.5, indicating that relatively more labor is employed by the second sector, because as the second stock of pollutant is less harmful the aggregate intertemporal damage of one unit of emissions is smaller for the second pollutant. This expectation is confirmed by the numerical optimization.

The third scenario ($\sigma_1 < \sigma_2, \delta_1 < \delta_2$) – both harmfulness and deterioration rates are different and the more harmful pollutant has the higher deterioration rate – is the most interesting as neither of the two pollutants exhibits a priori more favorable dynamic characteristics for the economy. Hence, we are not able to predict which production sector will be used to a greater extent in the stationary state. Furthermore, non monotonic paths – if they occur – are likely to be more pronounced than in the other cases. Figure 6.3 shows the optimal paths for a numerical example for all four subcases (initial pollutant stocks

[6]Note that both consumption goods are equally valued by the representative consumer, i.e. $\mu_1 = \mu_2$ (see Appendix A6.7).

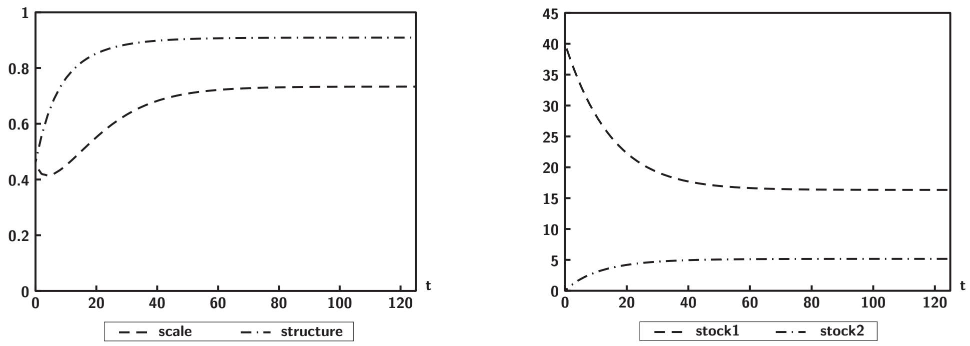

Figure 6.2 Optimal paths for scale and structure (left) and the two pollutant stocks (right) for the case $\sigma_1 < \sigma_2$, $\delta_1 = \delta_2$. Parameter values used for the numerical optimization are given in Appendix A6.7.

above or below stationary state levels for one and both pollutants). Of course, the long run stationary state to which the economy converges, is the same in all four subscenarios, as all parameters are identical except for the initial stocks of the two pollutants. Nevertheless, the optimal paths and especially their convergence towards the stationary state is quite different for the four subcases. As expected from Proposition 6.4, we observe that – if at all – the optimal path for the structure is non-monotonic if both stocks start above or below their stationary state levels (subcases a and d) and the optimal path for the scale is non-monotonic if one initial stock is higher and one is lower than their stationary state levels (subcase b). We also see that both, structure and scale, may exhibit monotonic optimal paths (subcase c).

In the fourth scenario ($\sigma_1 < \sigma_2, \delta_1 > \delta_2$), where both pollutants exhibit different harmfulness and deterioration rates but the second pollutant is more harmful and has the lower deterioration rate, the first pollutant exhibits clearly more favorable dynamic properties than the second pollutant. In this case the economy will nearly exclusively use the first production sector. Although non-monotonicities in the optimal paths for scale and structure can occur according to Proposition 6.4, they are not pronounced. As nothing new can be learned from this case, we do not show a numerical optimization example.

6.5 Conclusion

In this chapter, we have studied the mutual interaction over time between the scale and structure of economic activity on the one hand, and the dynamics of multiple environmental pollution stocks on the other hand. We have carried out a total analysis of a two-sector-economy, in which each sector produces one distinct consumption good and one specific pollutant. The pollutants of

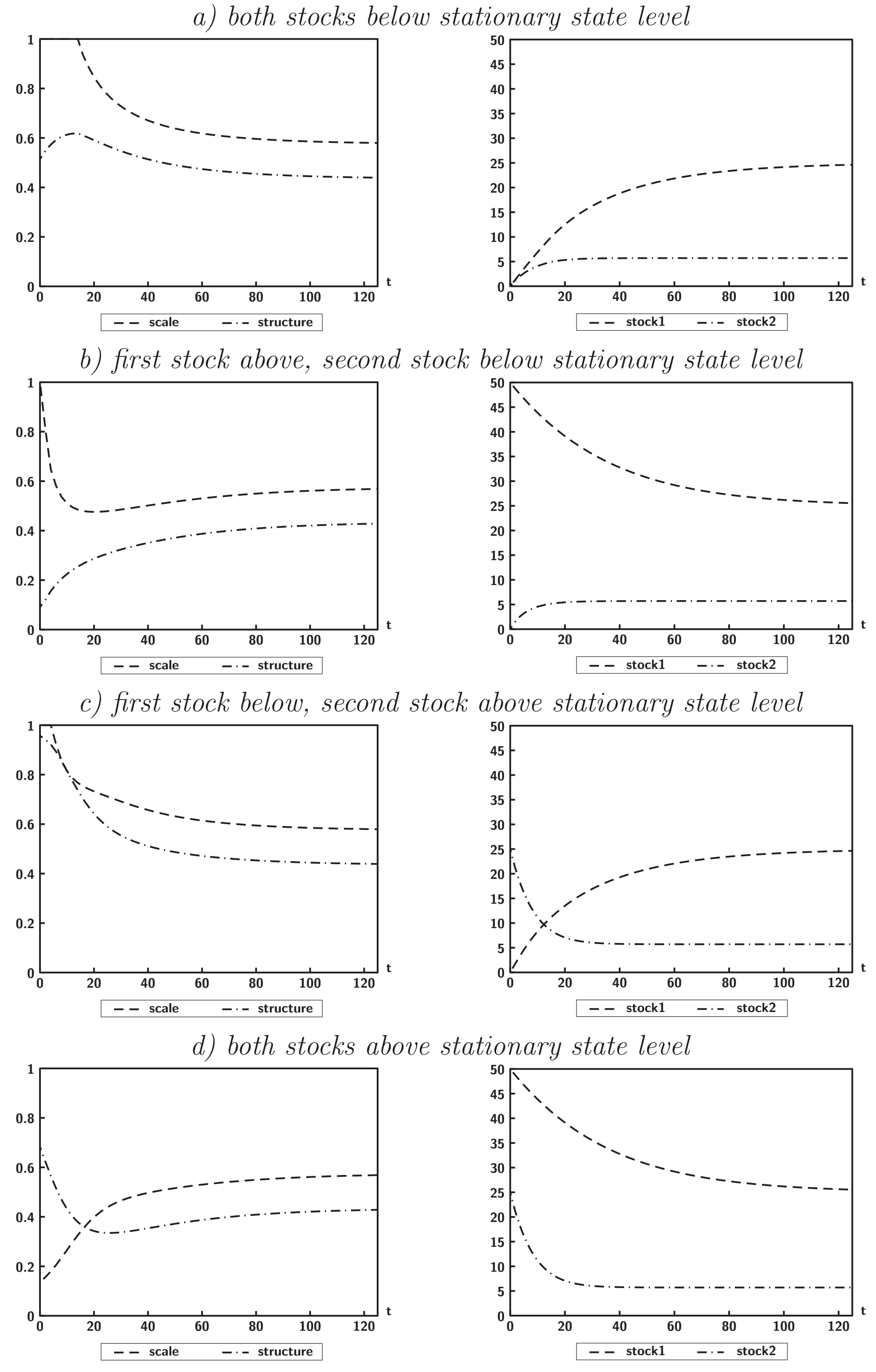

Figure 6.3 Optimal paths for scale and structure (left) and the two pollutant stocks (right) for the case $\sigma_1 < \sigma_2$, $\delta_1 < \delta_2$ and all four subscenarios. Parameter values used for the numerical optimization are given in Appendix A6.7.

both sectors were assumed to differ in their environmental impact in two ways: (i) with respect to their harmfulness and (ii) with respect to their natural deterioration rates in the environment.

Most of the results are intuitive. First, it may be optimal not to use all available labor endowment in the production of consumption goods in order to avoid excessive environmental damage. Second, under very general conditions a change in scale and structure of economic activity over time is optimal. Thus, the optimal economic dynamics is driven by the dynamics of the environmental pollution stocks. The less harmful is a pollutant, the higher are the relative importance of the emitting sector and the overall scale of economic activity in the stationary state. The shorter lived is a pollutant, the higher are the relative importance of the emitting sector and the overall scale of economic activity in the stationary state. If emissions differ either in their environmental harmfulness or in their deterioration rates, we should have structural change towards the sector emitting the less harmful or the shorter-lived pollutant. However, if the harmfulness and deterioration rates differ and if the environmentally less harmful emission is also the longer-lived pollutant, no general conclusion concerning the direction of structural change can be drawn. Third, the characteristic time scale of convergence of scale and structure towards the stationary state is given by (the inverse of) the eigenvalue with the smaller absolute value. It increases with the discount rate and the curvature of consumption welfare in the stationary state; it decreases with the harmfulness and the deterioration rate of the respective pollutant stock.

Most importantly, our formal analysis as well as the numerical optimizations, show that it is likely that the optimal control paths, i.e. the change in the scale and structure of the economy, are non-monotonic over time.[7] If a non-monotonic control is optimal, our numerical optimizations suggest that the local extremum of the control path may be pronounced and that it occurs at the beginning of the control path.

These results have implications for the design of environmental indicators and policies. First, the traditional view is that different environmental problems – such as e.g. acidification of soils and surface waters, groundwater contamination by nitrates or pesticides, and climate change due to anthropogenic greenhouse gas emission – can be regulated by independent environmental policies. In contrast, our total analysis of a multi-sector economy with several independent environmental pollutants, shows that these problems – even without any direct physical interaction – interact indirectly because they all affect social welfare, and the mitigation of all of them is constrained by the available eco-

---

[7]Non-monotonic optimal control paths, in particular limit-cycles, are known to exist for control problems with two or more state variables, and for time-lagged and adaptive control problems, even with one single state variable (e.g. Benhabib and Nishimura 1979, Feichtinger et al. 1994, Wirl 2000, 2002, Winkler 2004).

nomic resources. As a result, even for non-interacting environmental pollutants the optimal regulation has to take an encompassing view, taking into account all of the environmental problems together.

Second, indicators and policies which are solely based on the harmfulness of environmental pollutants – which is predominant in current environmental politics – fall short of optimally controlling environmental problems. In a dynamic setting, the lifetime of pollutants is an equally important determinant of the optimal environmental policy.

Third, the non-monotonicity-result challenges common intuition which suggests that policies should achieve optimal change in a monotonic way. In contrast to this simple intuition, our analysis shows that if pollutants accumulate on different time scales and if they differ in environmental harmfulness, the optimal policies may be non-monotonic. In particular, the optimal time-path of structural change towards the stationary state structure may be characterized by 'optimal overshooting'; that is, the optimal relative importance of a sector starts below (above) the stationary state level, increases (decreases) to a point above (below) the stationary state level, and finally decreases (increases) again. The same goes for the optimal dynamics of the overall economic scale.

Summing up, in order to develop sustainable solutions to the multiple environmental problems that we face in reality – such as climate change, depletion of the ozone layer, groundwater contamination, acidification of soil and surface water, biodiversity loss, etc. – we should adopt an encompassing view and base policy advice on a total analysis of economy-environment interactions. As our analysis shows, the resulting optimal policies need to take account of the history, the empirical parameter values and the dynamic relationships of all of the problems, and these policies might be non-monotonic.

Appendix

A6.1 Concavity of the Optimized Hamiltonian

We show that the Hamiltonian $\mathcal{H}$, without taking into account the restriction $c \leq 1$, i.e. $p_c = 0$, is strictly concave whenever the necessary conditions are satisfied. Thus, the unique optimal solution is the local extremum of $\mathcal{H}$ if we have an interior solution; it is a corner solution with $c = 1$ if the local extremum of $\mathcal{H}$ is reached for unfeasible $c > 1$.

A sufficient condition for strict concavity of the Hamiltonian is that its

Hessian $H = \frac{\partial^2 \mathcal{H}}{\partial i \partial j}$ $(i, j = c, x, s_1, s_2)$ is negative definite. The Hessian H reads:

$$H = \begin{pmatrix} \mathcal{H}_{cc} & \mathcal{H}_{cx} & 0 & 0 \\ \mathcal{H}_{xc} & \mathcal{H}_{xx} & 0 & 0 \\ 0 & 0 & -\sigma_1 & 0 \\ 0 & 0 & 0 & -\sigma_2 \end{pmatrix} \tag{A6.1}$$

Due to its diagonal form, H is negative definite if the reduced Hessian $H' = \frac{\partial^2 \mathcal{H}}{\partial i \partial j}$ $(i, j = c, x)$ is negative definite, i.e. $\mathcal{H}_{cc}, \mathcal{H}_{xx} < 0$ and $\det H' > 0$.

$$\mathcal{H}_{cc} = U_{11}(F_c^1)^2 + (U_1 + p_1)F_{cc}^1 + U_{22}(F_c^2)^2 + (U_2 + p_2)F_{cc}^2 \ , \tag{A6.2}$$
$$\mathcal{H}_{xx} = U_{11}(F_x^1)^2 + (U_1 + p_1)F_{xx}^1 + U_{22}(F_x^2)^2 + (U_2 + p_2)F_{xx}^2 \ , \tag{A6.3}$$
$$\mathcal{H}_{cx} = U_{11}F_c^1 F_x^1 + (U_1 + p_1)F_{cx}^1 + U_{22}F_c^2 F_x^2 + (U_2 + p_2)F_{cx}^2 \ . \tag{A6.4}$$

Along the optimal path, the necessary conditions have to be satisfied. In particular, for an interior solution, i.e. $c^\star < 1$, the necessary and sufficient conditions (6.19) and (6.20) become:

$$(U_1 + p_1)F_c^1 + (U_2 + p_2)F_c^2 = 0 \ , \tag{A6.5}$$
$$(U_1 + p_1)F_x^1 + (U_2 + p_2)F_x^2 = 0 \ . \tag{A6.6}$$

Thus, for an interior optimal path the following equations hold:

$$p_i = -U_i \qquad (i = 1, 2) \ . \tag{A6.7}$$

With this, one obtains:

$$\mathcal{H}_{cc} = U_{11}(F_c^1)^2 + U_{22}(F_c^2)^2 < 0 \ , \tag{A6.8}$$
$$\mathcal{H}_{xx} = U_{11}(F_x^1)^2 + U_{22}(F_x^2)^2 < 0 \ , \tag{A6.9}$$
$$\begin{aligned} \det H' &= \mathcal{H}_{cc}\mathcal{H}_{xx} - \mathcal{H}_{cx}^2 \\ &= U_{11}U_{22}\left[(F_c^1)^2(F_x^2)^2 + (F_x^1)^2(F_c^2)^2 - 2F_c^1 F_x^1 F_c^2 F_x^2\right] > 0 \end{aligned} \tag{A6.10}$$

Hence, whenever $\mathcal{H}$ has an extremum it is a maximum. As a consequence, the necessary conditions (plus the transversality condition 6.24) are also sufficient.

A6.2 Proof of Proposition 6.1

(i) Inserting Equations (6.29) into Equations (6.26) and (6.27), and using the relationship between F^i and P^i, as given from Equation (6.9), one obtains:

$$U_i^\star = \frac{\sigma_i P^{i\star}}{\delta_i(\delta_i + \rho)} + \frac{p_c}{\lambda P_l^{i\star}} \qquad (i = 1, 2) \ . \tag{A6.11}$$

With the properties for P^i, as given by (6.1), and the properties for U_i, as given by (6.7), the left-hand-side of Equation (A6.11) is strictly decreasing

while the right-hand-side is strictly increasing in l_i. Thus, there exists at most one $l_i^\star$ which satisfies Equation (A6.11). The existence of such a solution is guaranteed by the properties $\lim_{l_i \to 0} P_l^i = +\infty$ and $\lim_{y_i \to 0} U_i = +\infty$.
(ii) We derive $\bar{l}_i$ by solving (A6.11) for $l_i^\star$ assuming $p_c = 0$. Thus, $\bar{l}_i$ is the maximal amount of labor which will be assigned to production process i in an optimal stationary state without taking account for the restriction $c \leq 1$. If $\bar{l}_1 + \bar{l}_2 \geq \lambda$ the labor supply is short of the optimal labor demand and thus the stationary state is a corner solution. If, on the other hand, the total labor supply λ exceeds the sum $\bar{l}_1 + \bar{l}_2$, then not all labor will be used for economic activity and the optimal stationary state will be an interior solution.

A6.3 Proof of Proposition 6.2

Setting $p_c = 0$ in Equation (A6.11) yields for an interior stationary path:

$$U_i^\star = \frac{\sigma_i F^{i\star}}{\delta_i(\delta_i + \rho)} \qquad (i = 1, 2) \ . \tag{A6.12}$$

By implicit differentiation of (A6.12) with respect to δ_j $(j = 1, 2)$ one obtains:

$$\begin{aligned}
\left(F_c^{j\star}\frac{\partial c^\star}{\partial \delta_j} + F_x^{j\star}\frac{\partial x^\star}{\partial \delta_j}\right)\left(U_{jj}^\star - \frac{\sigma_j}{\delta_j(\delta_j+\rho)}\right) &= -\frac{\sigma_j F^{j\star}(2\delta_j + \rho)}{\delta_j^2(\delta_j+\rho)^2} \qquad (j = i) \ , \\
\left(F_c^{i\star}\frac{\partial c^\star}{\partial \delta_j} + F_x^{i\star}\frac{\partial x^\star}{\partial \delta_j}\right)\left(U_{ii}^\star - \frac{\sigma_i}{\delta_i(\delta_i+\rho)}\right) &= 0 \qquad (j \neq i) \ .
\end{aligned}$$

Solving for $\partial c^\star/\partial \delta_j$ and $\partial x^\star/\partial \delta_j$ yields:

$$\frac{\partial c^\star}{\partial \delta_j} = \frac{\sigma_j F^{j\star} F_x^{i\star}(2\delta_j + \rho)}{(F_c^{i\star}F_x^{j\star} - F_c^{j\star}F_x^{i\star})(U_{jj}^\star \delta_j(\delta_j+\rho) - \sigma_j)\delta_j(\delta_j+\rho)} \ , \tag{A6.13}$$

$$\frac{\partial x^\star}{\partial \delta_j} = \frac{\sigma_j F^{j\star} F_c^{i\star}(2\delta_j + \rho)}{(F_c^{j\star}F_x^{i\star} - F_c^{i\star}F_x^{j\star})(U_{jj}^\star \delta_j(\delta_j+\rho) - \sigma_j)\delta_j(\delta_j+\rho)} \ . \tag{A6.14}$$

From the signs of the F_j^i $(i = 1, 2; j = c, x)$ it follows that

$$\frac{\partial c^\star}{\partial \delta_1} > 0 \ , \qquad \frac{\partial c^\star}{\partial \delta_2} > 0, \qquad \frac{\partial x^\star}{\partial \delta_1} > 0, \qquad \frac{\partial x^\star}{\partial \delta_2} < 0 \ . \tag{A6.15}$$

By implicit differentiation of (A6.12) with respect to σ_j $(j = 1, 2)$ one obtains:

$$\begin{aligned}
\left(F_c^{j\star}\frac{\partial c^\star}{\partial \sigma_j} + F_x^{j\star}\frac{\partial x^\star}{\partial \sigma_j}\right)\left(U_{jj}^\star - \frac{\sigma_j}{\delta_j(\delta_j+\rho)}\right) &= \frac{F^{j\star}}{\delta_j(\delta_j+\rho)} \qquad (j = i) \ , \\
\left(F_c^{i\star}\frac{\partial c^\star}{\partial \sigma_j} + F_x^{i\star}\frac{\partial x^\star}{\partial \sigma_j}\right)\left(U_{ii}^\star - \frac{\sigma_i}{\delta_i(\delta_i+\rho)}\right) &= 0 \qquad (j \neq i) \ .
\end{aligned}$$

Solving for $\partial c^\star/\partial\sigma_j$ and $\partial x^\star/\partial\sigma_j$ yields:

$$\frac{\partial c^\star}{\partial\sigma_j} = \frac{F^{j\star}F_x^{i\star}}{(F_c^{j\star}F_x^{i\star}-F_c^{i\star}F_x^{j\star})(U_{jj}^\star\delta_j(\delta_j+\rho)-\sigma_j)}\ , \tag{A6.16}$$

$$\frac{\partial x^\star}{\partial\sigma_j} = \frac{F^{j\star}F_c^{i\star}}{(F_c^{i\star}F_x^{j\star}-F_c^{j\star}F_x^{i\star})(U_{jj}^\star\delta_j(\delta_j+\rho)-\sigma_j)}\ . \tag{A6.17}$$

From the signs of the F_j^i $(i=1,2; j=c,x)$ it follows that

$$\frac{\partial c^\star}{\partial\sigma_1}<0\ ,\qquad \frac{\partial c^\star}{\partial\sigma_2}<0,\qquad \frac{\partial x^\star}{\partial\sigma_1}<0,\qquad \frac{\partial x^\star}{\partial\sigma_2}>0\ . \tag{A6.18}$$

Implicit differentiation of (A6.12) with respect to ρ yields:

$$F_c^{1\star}\frac{\partial c^\star}{\partial\rho}+F_x^{1\star}\frac{\partial x^\star}{\partial\rho} = -\frac{\sigma_1F^{1\star}}{[U_{11}^\star\delta_1(\delta_1+\rho)-\sigma_1](\delta_1+\rho)}\ ,$$

$$F_c^{2\star}\frac{\partial c^\star}{\partial\rho}+F_x^{2\star}\frac{\partial x^\star}{\partial\rho} = -\frac{\sigma_2F^{2\star}}{[U_{22}^\star\delta_2(\delta_2+\rho)-\sigma_2](\delta_2+\rho)}\ ,$$

Solving for $\partial c^\star/\partial\rho$ and $\partial x^\star/\partial\rho$ yields:

$$\frac{\partial c^\star}{\partial\rho}(F_c^{2\star}F_x^{1\star}-F_c^{1\star}F_x^{2\star}) =$$

$$= \frac{\sigma_1F^{1\star}F_x^{2\star}}{[U_{11}^\star\delta_1(\delta_1+\rho)-\sigma_1](\delta_1+\rho)}-\frac{\sigma_2F^{2\star}F_x^{1\star}}{[U_{22}^\star\delta_2(\delta_2+\rho)-\sigma_2](\delta_2+\rho)}, \tag{A6.19}$$

$$\frac{\partial x^\star}{\partial\rho}(F_c^{2\star}F_x^{1\star}-F_c^{1\star}F_x^{2\star}) =$$

$$= \frac{\sigma_2F^{2\star}F_c^{1\star}}{[U_{22}^\star\delta_2(\delta_2+\rho)-\sigma_2](\delta_2+\rho)}-\frac{\sigma_1F^{1\star}F_c^{2\star}}{[U_{11}^\star\delta_1(\delta_1+\rho)-\sigma_1](\delta_1+\rho)}. \tag{A6.20}$$

From the signs of the F_j^i $(i=1,2; j=c,x)$ it follows that

$$\frac{\partial c^\star}{\partial\rho}>0\ ,\qquad \frac{\partial x^\star}{\partial\rho}\gtreqless 0 \Leftrightarrow \frac{[U_{22}^\star\delta_2(\delta_2+\rho)-\sigma_2](\delta_2+\rho)}{[U_{11}^\star\delta_1(\delta_1+\rho)-\sigma_1](\delta_1+\rho)}\gtreqless\frac{\sigma_2F^{2\star}F_c^{1\star}}{\sigma_1F^{1\star}F_c^{2\star}}\ . \tag{A6.21}$$

A6.4 Derivation of the Differential Equation System

Differentiation of $p_i=-U_i$ (Equation A6.7) with respect to time and inserting into Equations (6.21) and (6.22) yields, together with the equations of motion (6.15), a system of four differential equations in the four unknowns c, x, s_1 and s_2:

$$\sigma_1s_1-U_1(\delta_1+\rho)+U_{11}(F_c^1\dot{c}+F_x^1\dot{x}) = 0\ , \tag{A6.22}$$

$$\sigma_2s_2-U_2(\delta_2+\rho)+U_{22}(F_c^2\dot{c}+F_x^2\dot{x}) = 0\ , \tag{A6.23}$$

$$\dot{s_1}-F^1+\delta_1s_1 = 0\ , \tag{A6.24}$$

$$\dot{s_2}-F^2+\delta_2s_2 = 0\ . \tag{A6.25}$$

The conditions (A6.22)–(A6.25) for an interior optimal solution can be rearranged to yield the system (6.30)–(6.33) of four coupled autonomous differential equations.

A6.5 Eigenvalues and Eigenvectors of the Jacobian

We obtain the Jacobian J^* by differentiating the right-hand-sides of Equations (6.30)–(6.33) with respect to c, x, s_1 and s_2 and evaluating them at the stationary state. Taking into account that in the interior stationary state (6.28) holds with equality, $U_i = \sigma_i s_i^\star/(\delta_i + \rho)$, one obtains for the Jacobian $J^\star$:

$$J^* = \begin{pmatrix} \rho + \frac{\delta_1 F_c^{1\star} F_x^{2\star} - \delta_2 F_x^{1\star} F_c^{2\star}}{df^\star} & \frac{(\delta_1-\delta_2) F_x^{1\star} F_x^{2\star}}{df^\star} & -\frac{\sigma_1 F_x^{2\star}}{U_{11}^\star df^\star} & \frac{\sigma_2 F_x^{1\star}}{U_{22}^\star df^\star} \\ \frac{(\delta_2-\delta_1) F_c^{1\star} F_c^{2\star}}{df^\star} & \rho + \frac{\delta_2 F_c^{1\star} F_x^{2\star} - \delta_1 F_x^{1\star} F_c^{2\star}}{df^\star} & \frac{\sigma_1 F_c^{2\star}}{U_{11}^\star df^\star} & -\frac{\sigma_2 F_c^{1\star}}{U_{22}^\star df^\star} \\ F_c^{1\star} & F_x^{1\star} & -\delta_1 & 0 \\ F_c^{2\star} & F_x^{2\star} & 0 & -\delta_2 \end{pmatrix} .$$

The eigenvalues ν_i and eigenvectors ξ_i are the solutions of the equation $J^\star \cdot \xi = \nu \cdot \xi$. The four eigenvalues are:

$$\nu_1 = \frac{1}{2}\left[\rho - \sqrt{(\rho + 2\delta_1)^2 - \frac{4\sigma_1}{U_{11}^\star}}\right] < 0 \, , \tag{A6.26}$$

$$\nu_2 = \frac{1}{2}\left[\rho - \sqrt{(\rho + 2\delta_2)^2 - \frac{4\sigma_2}{U_{22}^\star}}\right] < 0 \, , \tag{A6.27}$$

$$\nu_3 = \frac{1}{2}\left[\rho + \sqrt{(\rho + 2\delta_1)^2 - \frac{4\sigma_1}{U_{11}^\star}}\right] > 0 \, , \tag{A6.28}$$

$$\nu_4 = \frac{1}{2}\left[\rho + \sqrt{(\rho + 2\delta_2)^2 - \frac{4\sigma_2}{U_{22}^\star}}\right] > 0 \, . \tag{A6.29}$$

The eigenvectors associated with the negative eigenvalues ν_1 and ν_2 are:

$$\xi_1 = \left(\frac{F_x^{2\star}(\nu_1 + \delta_1)}{df^\star}, -\frac{F_c^{2\star}(\nu_1 + \delta_1)}{df^\star}, 1, 0\right) \, , \tag{A6.30}$$

$$\xi_2 = \left(-\frac{F_x^{1\star}(\nu_2 + \delta_2)}{df^\star}, \frac{F_c^{1\star}(\nu_2 + \delta_2)}{df^\star}, 0, 1\right) \, . \tag{A6.31}$$

A6.6 Time Scale of Convergence

Equations (6.35) and (6.36) are of the following type:

$$z(t) = z^\star + Ae^{\nu_1 t} + Be^{\nu_2 t} \qquad (\nu_1, \nu_2 < 0) \, , \tag{A6.32}$$

with real constants A and B. Without loss of generality assume that $|\nu_1| < |\nu_2|$. Since we are interested in the system dynamics in a neighborhood of the stationary state, we calculate the characteristic time scale of convergence for z as $t \to \infty$. According to (6.39), the characteristic time scale of convergence of z in a neighborhood of the stationary state $z^\star$ is given by:

$$\tau_z^{-1} = \overline{\left| \lim_{t\to\infty} \frac{A\nu_1 e^{\nu_1 t} + B\nu_2 e^{\nu_2 t}}{Ae^{\nu_1 t} + Be^{\nu_2 t}} \right|} = \overline{\left| \lim_{t\to\infty} \frac{A\nu_1 + B\nu_2 e^{(\nu_2-\nu_1)t}}{A + Be^{(\nu_2-\nu_1)t}} \right|} = |\nu_1| \ . \quad \text{(A6.33)}$$

Hence, for $t \to \infty$ the characteristic time scale of convergence is constant and given by $1/\min\{|\nu_1|, |\nu_2|\}$.

A6.7 Parameter Values for the Numerical Optimization

We used a Cobb-Douglas welfare function for the numerical optimizations,

$$U(y_1, y_2) = 0.5\ln(y_1) + 0.5\ln(y_2) \ , \quad \text{(A6.34)}$$

and the following production functions:

$$P^1(l_1) = \sqrt{l_1} \ , \qquad P^2(l_2) = \sqrt{l_2} \ . \quad \text{(A6.35)}$$

For all numerical optimizations we set $\lambda = 1$ and $\rho = 0.03$. In addition, we used the following parameter values for the different scenarios:

Figure	σ_1	σ_2	δ_1	δ_2	s_1	s_2
1	0.01	0.01	0.02	0.1	30	30
2	0.003	0.03	0.05	0.05	40	0
3a	0.002	0.02	0.02	0.1	0	0
3b	0.002	0.02	0.02	0.1	50	0
3c	0.002	0.02	0.02	0.1	0	25
3d	0.002	0.02	0.02	0.1	50	25

PART II

Biodiversity

7. Biodiversity as an Economic Good*

7.1 Introduction

Biological diversity, which has been defined as 'the variability among living organisms from all sources [...] and the ecological complexes of which they are part' (CBD 1992), is currently being lost at rates that exceed the natural extinction rates of the past by a factor of somewhere between 100 and 1,000 (Watson et al. 1995b). This is one of the most eminent environmental problems of our time (Wilson 1988).

Ecologists were among the first to point out this alarming development and to express concern over its potential negative effect on ecosystems and human well-being (Ehrlich and Ehrlich 1981, Myers 1979, Soulé 1986, Soulé and Wilcox 1980, Wilson 1988). They vindicated their concern with the important role that biodiversity plays for ecosystem functioning (Holling et al. 1995, Loreau et al. 2001, Schulze and Mooney 1993, Tilman 1997a) and for providing essential life-support services to the human existence on planet Earth (Daily 1997b, Mooney and Ehrlich 1997, Perrings et al. 1995b). Examples of such ecosystem services include nutrient cycling, control of water runoff, purification of air and water, soil regeneration, pollination of crops and natural vegetation, or climate stabilization. The main mechanisms through which biodiversity is currently being lost were identified to be loss of habitat, overuse of populations, invasion of non-native species, pollution of ecosystems and climate change (Barbier et al. 1994, Watson et al. 1995a).

Economists have pointed out the high economic value of biodiversity, which comprises both use and non-use values (Goulder and Kennedy 1997, Randall 1988, Watson et al. 1995a). Many species have direct use value as food, fuel, construction material, industrial resources or pharmaceutical substances (Farnsworth 1988, Plotkin 1988). Beyond that, biodiversity, i.e. the set of all species, has an important indirect use value in so far as entire ecosystems per-

*Translated and revised from 'Der ökonomische Wert der biologischen Vielfalt', in: Bayerische Akademie für Naturschutz und Landschaftspflege (ed.), *Grundlagen zum Verständnis der Artenvielfalt und ihrer Bedeutung und der Maßnahmen, dem Artensterben entgegen zu wirken* (Laufener Seminarbeiträge 2/02), Laufen/Salzach, 2002, pp. 73–90.

form valuable services as described above. One study went so far as to estimate the total economic value of all the Planet's ecosystem goods and services at US$ 33 trillion per year, a number comparable in order of magnitude to aggregate world GDP (Costanza et al. 1997b). Besides putting a value on what is currently being lost and what is at risk, economists also provided a number of explanations for the loss of biodiversity currently being observed (Barbier et al. 1994, Watson et al. 1995a, Moran and Pearce 1997). The fundamental causes of biodiversity loss include the growth of human population; market failure because of externalities and the public good character of biological resources; governance failure in regulating the access to, and use of, biological resources; and fundamental ignorance pertaining to both individual and social decision making.

By now, the international community has acknowledged the problem of biodiversity loss, and the need to enact policies to halt or even reverse this problem. In June 1992, the *Convention on Biological Diversity* was signed by 156 states at the United Nations Conference on Environment and Development in Rio de Janeiro, Brazil (CBD 1992). In the preamble of this convention, the signatories explicitly declare that biodiversity has – besides ecological, cultural, spiritual and intrinsic values – also an economic value.

Yet, many ecologists (and even more environmentalists) regard the contribution of economists to the discussion about biodiversity loss and conservation with great suspicion: 'Isn't it the economy, which causes biodiversity loss? And don't economists always argue in favor of economic interests?' While this suspicion is largely based on ignorance about the nature and substance of economics as a science, there is a corresponding reluctance among professional economists to engage in the discussion about biodiversity loss and conservation: 'What exactly is this thing called "biodiversity"? In what sense is it an economic good? And how can we discuss its efficient and fair allocation based on standard economic concepts?' For example, Weimann et al. (2003: 7) express this wide-spread unease concerning biodiversity among economists, by pointing out that '[b]iodiversity exists, but there is no consensus about what it is. Biodiversity is finite and declining, but there is no consensus about how to measure it. Biodiversity is important, but there is no consensus about how important it is, and for whom' (own translation).

The discussion of biodiversity loss and conservation therefore faces the following challenge. On the one hand, economics seems to have important contributions to make to this discussion. But on the other hand, a number of conceptual questions remain open, which need to be addressed before economists can apply their standard tools and methods in an analysis:

(i) In what sense can one think of biodiversity as an economic good?

(ii) How can one quantitatively measure biodiversity?

(iii) In what sense does biodiversity have economic value?

In this and the following chapter, I shall address these questions with the aim of clarifying the conceptual foundations upon which an ecological-economic analysis of biodiversity loss and conservation is possible. In this chapter, questions (i) and (iii) are addressed. It is discussed in what sense one can think of biological diversity as an economic good, and what constitutes its economic value. This is not to neglect or belittle the importance of the other value dimensions in any way. Rather, it will be shown that considering the economic value of biological diversity can yield important insights for an encompassing understanding of biodiversity loss and conservation. Question (ii), of how to quantitatively measure biodiversity, will then be addressed in Chapter 8. In that chapter, I will also discuss the intricate relationship between the measurement and the valuation of biodiversity.

The argument in this chapter proceeds as follows. In Section 7.2, it is discussed whether, and to what extent, one can consider biodiversity as an economic good. On this basis, one can then specify what its economic value is (Section 7.3). Based on the economic value of biodiversity, one can explain the current loss of biodiversity from an economic perspective, and identify its fundamental causes (Section 7.4). The economic value of biodiversity also offers a conceptual framework for discussing the question 'What species and populations should be protected, and to what extent?' on scientific grounds (Section 7.5). This allows one to prioritize different biodiversity protection goals. In conclusion, viewing biodiversity as an economic good which has economic value, makes apparent the potential and limits of economics as an academic discipline for the analysis of biodiversity loss and conservation (Section 7.6).

7.2 Biological Diversity as an Economic Good

According to a classic definition, 'economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses' (Robbins 1932: 15). In this sense, biological diversity can be thought of as an economic good (Heal 2000). It satisfies human needs and allows people to achieve certain ends in a variety of ways. On the other hand, biodiversity is scarce and can be used in alternative ways. Both aspects will be explained in detail in the following.

7.2.1 Satisfaction of Human Needs

Biological diversity, and its components, can satisfy human needs and allow people to achieve certain ends in a variety of ways. The following examples may illustrate this claim.

Food

A large part of our current food supply comes from domesticated plant and animal species, which have originally been derived from wild species. Of the 240,000 known (vascular) plant species an estimated 25% are edible (Watson et al. 1995b: 13), that is, 60,000. In the course of human history only 3,000 species of these have ever served as food, only 150 species have ever been cultivated on a larger scale, and today less than 20 satisfy more than 90% of total human food demand (Myers 1989: 54). The largest share is made up by only four species – wheat, corn, rice and potato – which cover more than 50% of the whole demand for vegetable food (Plotkin 1988: 107).

Besides specializing on fewer and fewer species, the genetic diversity of edible plants and animals is being diminished also within individual species, by using only a few high yield varieties per species. These are being selected by breeding with respect to certain preferred properties, in particular large and homogenous amounts of raw product. As a result, in many countries in which traditionally a large diversity of different varieties have been grown, today only very few are still being cultivated. For example, the number of rice varieties grown in Sri Lanka has decreased from 2,000 in 1959 to currently only five (Swanson 1994: 26f.).

On the one hand, this process of specialization leads to significantly higher average yields per hectare. But on the other hand, it is accompanied by an increased susceptibility to diseases, pests or extreme weather conditions. In order to avoid negative consequences due to these susceptibilities, and in order to further increase yields for the food demand of a growing world population, modern agriculture necessarily depends on crossbreeding with genetic material from wild varieties, which is available in natural ecosystems. These species develop under largely natural condition and, hence, can permanently develop new defensive mechanisms against pests and diseases (Ehrlich and Ehrlich 1981: 65). At the same time, they provide the genetic raw material for other desired properties. For example, the properties of so-called halophytes – plants which are tolerant against salt – may be transferred to conventional species, which would mean an enormous gain in the potential area of cultivable land as well as the potential of irrigating with saltwater (Myers 1983: 54). Wild species in natural ecosystems therefore provide a reservoir of genetic diversity which is important for securing long-term food supplies (Heal et al. 2004).

Pharmaceuticals

Biological diversity makes an important contribution to the supply of humankind with pharmaceutical substances. Its particular utility in this respect stems from the fact that the different organisms in their biotic environment have developed a number of survival strategies, by developing biologically active chemical substances which have proved successful in the course of evolution. These chemical substances may also be useful for humans in many instances because humans have to survive in the same natural system and in interaction with the same other life forms (Swanson 1996: 3). Already today, humans depend to a large extent on wild organisms in their supply with pharmaceutical substances. Myers (1997: 263) estimates that one quarter of all registered pharmaceutical substances stem from plants, another quarter stem from animals and microorganisms.

One can distinguish between three different approaches of how plant or animal species are being used by the pharmaceutical industry (Swanson et al. 1992: 434). First, parts which have been isolated from plants or animals may be directly used as a therapeutic substance. For example, one can isolate different substances from snake poison, which inhibit or enhance the coagulation of blood and may be used for the regulation and diagnostics of various blood diseases (Hall 1992: 380). Second, parts of plants or animals may be used as a raw material in the process of synthesizing pharmaceutical substances. Third, parts of plants or animals can serve as an exemplar for designing and synthesizing pharmaceutical substances in the lab. The most well-known example is aspirin, which was originally produced from the leaves of the willow tree, but can be produced today at much lower costs synthetically.

In 1993, roughly 80% of the 150 most prescribed drugs in the USA were synthetic drugs which were designed after the exemplar of natural substances, half-synthetic substances from natural substances or, in a few cases, natural substances (Watson et al. 1995b: 14), and the worldwide sales of pharmaceuticals on the basis of plant substances was worth 59 billion US dollars (ten Kate 1995).[1] These successful pharmaceutical substances have been identified although only 5,000 of the estimated 240,000 vascular plants have been researched systematically and thoroughly for their potential as a pharmaceutical substance (Oldfield 1992: 350). Obviously, biological diversity offers considerable potential for the development of new pharmaceutical substances. This potential is currently subject to great commercial as well as economic interest and is being targeted by so-called 'bio-prospecting' (Mateo et al. 2000, Polasky and Solow 1995, Polasky et al. 1993, Rausser and Small 2000, Simpson et al. 1996).

---

[1]The top three pharmaceutical substances in terms of sales, from plants, animals and microorganisms in 1997, were (WMPQ 1999): (1) Zocor (sales: 3.6 billion US dollars), a cholesterol-synthesis-inhibitor from Merck & Co., which is produced after the exemplar of the natural agent lovastatin from the fungus *Aspergillus terrestris*; (2) Vasotec (sales: 2.5 billion US dollars), an ACE-inhibitor, also from Merck & Co., which was developed from a peptide in the poison of the fer-de-lance (*Bothrops jararaca* or *athrox*); (3) Augmentin (sales: 1.5 billion US dollars), from Smith-Kline-Beecham, with the agent co-amoxiclav, which is a combination of a beta-lactamase-inhibitor from the bacterium *streptomyces lavuligerus* and the half-synthetic antibiotic amoxicillin (*penicillium spp.* or *aspergillus spp.*).

Industrial resources

Biodiversity makes an important contribution to human welfare in its function as a supplier of industrial resources, which becomes more and more important as non-renewable resources (for example fossil fuels and mineral ores) become scarcer and scarcer. Different kinds of wood, rattan, rubber, fat, oil, wax, resin, vegetable dye, fibre, and many other resources are extracted from living organisms and are being used in many instances (Myers 1983: 146ff.). Biological diversity constitutes a stock of additional promising substances which may be used as industrial resources in the future. In particular, the chemical industry is increasingly interested in substances from living organisms. According to some estimates, this industry obtains more than 10% of its resources from agriculture and forestry already (Mann 1998: 60). The most important resource is still crude oil; but in light of the finite supply of fossil fuels, the substitution of this resource by plant resources is expected to become more and more important for the chemical industry (Myers 1983: 147).

Bioindicators for science

Biological diversity plays an important role as a source of new scientific insights and as a research model for science. For example, many species can inform medicinal research about the origin and nature of different human diseases (Myers 1983: 120). Research about hemophilia (bleeding disorder), for example, has been informed by the study of manatee (*dugong*), which have blood with bad coagulation properties. The armadillo (family *dasypodidae*) and the mangabey are the only species – besides humans – which can contract leprosy (Hansen's disease) and, hence, can yield important insights for research on this disease.

A special discipline – bionics – exists, which is concerned with systematically transferring problem solutions which have been developed and optimized over millions of years in nature into the technical domain (Hill 1997, Nader and Hill 1999). Engineers, for example, have gained insights for aircraft construction from studying the biophysical properties of insects and birds. Here, biological diversity serves as a role model for technical solutions.

A further use of biological diversity is bio-indication, which is the detection and quantification of anthropogenic environmental change by measuring changes in organisms and ecosystems (Arndt et al. 1987: 16). Bio-indication allows one to detect the existence of pollutants in different environmental media (for example air, soil, water), which is possible with technical devices only at much higher complexity and costs (Hampicke 1991: 30). For example, the heavy metal content of the atmosphere can be estimated based on the enrichment of heavy metal in mosses (Arndt et al. 1987: 57ff.), and algae can be used as indicators for the loading of aquatic ecosystems with organic substances and heavy metals (Arndt et al. 1987: 277ff.).

Aesthetical satisfaction and recreation

Biodiversity also satisfy human needs under aesthetic criteria. The beauty of many birds, butterflies, tropical fishes or flowering plants is beyond question and is certainly capable of satisfying the human need for aesthetic stimulation and contemplation. This is illustrated by a variety of different leisure activities, such as nature photography, bird watching, collection of butterflies, or diving (Ehrlich and Ehrlich 1992: 220). Even little and inconspicuous species are capable of fascinating the observer by particular properties, their complexity or unusual behavior. In this respect, it is exactly the diversity and the differences between species which matter (Ehrlich and Ehrlich 1981: 42).

As an indicator for the actual appreciation of biological diversity in its aesthetic and recreational function, one can take the increasing expenditures for eco-tourism. In 1988 approximately 235 million people worldwide took part in activities of eco-tourism, creating sales of an estimated 233 billion US dollars (Watson et al. 1995b: 16).

Ecosystem services

Ecosystems generate a number of functions and processes which ultimately satisfy human needs of consumption and production. The whole range of these so-called 'ecosystem services' (Daily 1997a: 3) can be classified into three main categories.

First, ecosystem services support human productive activities. For example, different species contribute decisively to the formation of soils, the conservation of the soil's fertility, and the protection against soil erosion. Thus, they fulfill important functions for agriculture and forestry. Furthermore, different species of microorganisms transform the nutrients in the soil (e.g. nitrogen, sulfur, phosphor etc.) into a form in which they can be processed by higher plants. These plants then carry out 'primary production', that is, by photosynthesis they transform the energy inflow from the sun into energy stored in chemical compounds, which can then be used as an energy source by animals. Agriculture also benefits from the control of the vast majority of agricultural pests by their natural enemies (Naylor and Ehrlich 1997: 151ff.), as well as of the pollination of agriculturally cultivated and wild flowering plants (Nabhan and Buchmann 1997: 133ff.).

Second, ecosystems serve as a sink for different wastes of human consumption and production. These are taken up, transformed and, thus, made partially innoxious or even reusable (Munasinghe 1992: 228). For example, the destruents in the soil decompose organic wastes into simpler inorganic components which can then serve again as nutrients for plants. Also, the bacteria in aquatic ecosystems are important destruents whose capability of decomposing wastes is being used today in sewage plants (Ehrlich and Ehrlich 1992: 222). Finally, the living parts in ecosystems also contribute to the decomposition of pesticides and air pollutants (McNeely et al. 1990: 32).

Third, ecosystems fulfill essential and irreplaceable life-support functions without which life on Earth could not exist in its present form (Munasinghe 1992: 228). Among these life supporting ecosystem services are the control of the gaseous composition of the atmosphere (oxygen,nitrogen and carbon dioxide content;[2] existence of the ozone layer which protects from UV radiation), the transformation ('primary production') of solar energy in biomass, in which form it can be used in the food chain by living beings who do not photosynthesize, regulation of water runoff in watersheds and general water circulation, regulation of local, regional and global climate, and regulation of nutrient cycles (carbon, oxygen, nitrogen, sulfur, phosphor, etc.) (Ehrlich und Ehrlich 1981: 86, Ehrlich and Ehrlich 1992: 221f., McNeely et al. 1990: 32).

The role of biodiversity for the capability of the ecosystems to generate all these services and to maintain their functioning, even under environmental changes, is still subject to scientific research. On the one hand, there are species the importance of which for the functioning of ecosystems exceeds by far their relative abundance in the ecosystem, for example the mykorrhiza fungi for the uptake of nutrients from the soil by plants (van der Heijden et al. 1998).[3] The loss of these so-called 'keystone species' (Bond 1993: 237ff.) would necessarily entail the loss of further species and strongly reduce the functional integrity of an ecosystem. In contrast, other species are highly redundant in the functions which they fulfill within the community (Lawton and Brown 1993). In the literature, these species are often called 'passenger species' (Holling et al. 1995: 67). The loss of one of these species can be compensated by another one (Watson et al. 1995a: 289). According to what we know today, at least in the short run, a small number of keystone species and physical processes suffice to guarantee the full functioning of ecosystems (Holling et al. 1995: 67). However, in the course of time, with changing environmental conditions, species which are currently passenger species can evolve into keystone species and overtake

[2]The oxygen content of the atmosphere has been constant at around 21% for the last approximately 350 million years due to the existence of green plants (Heintz and Reinhard 1993: 11–16). This is not only important as an essential component of the 'air to breathe'. A decrease in this fraction to 15% would imply that even dry wood could not burn any more, while an increase of this fraction to 25% would imply that even wet tropical forests could catch fire. This would have far-reaching implications for the development of ecosystems.

[3]*Mykorrhiza* ('fungus root') denotes the symbiosis between plants and soil fungi (Strasburger 1991: 229). The fungus fibre penetrates the plant's roots, such that an exchange of matter is possible. The plant thus uses the enormous absorptive capability of the fungi in order to obtain water and nutrients. Conversely, the fungus obtains sugar and carbohydrates from the plant, which the plant usually has in excess.

important functions within ecosystems (Barbier et al. 1994: 28). The functional diversity of species thus contributes to the resilience of ecosystems, that is, their ability to maintain ecosystem functions under changing environmental conditions (McCann 2000, Lehman and Tilman 2000).

Summary: Satisfaction of human needs

The examples listed here demonstrate the vast potential for utilizing biodiversity. The large differences between the examples also suggest that there is no universal criterion according to which one could make an à priori assessment of which components of biological diversity are of utility for humans and which are not (Hampicke 1991: 28). While in the past the economic relevance of the direct consequences of a loss of biological diversity for human consumption and production have been stressed; the focus, even in the economic research on biodiversity, is increasingly on the role that the loss of biodiversity has for the functioning and resilience of ecosystems (Barbier et al. 1994: 17, Perrings 1995c, Perrings et al. 1995b).

7.2.2 Scarcity

What makes biodiversity an economic good is, besides its economic utility, its scarcity (Lerch 1995: 33). Scarcity means that the provision or conservation of biodiversity is costly.[4] These costs can be monetary expenditures; for example, for the set up of a nature protection area. The financial resources, which may be spent on biodiversity conservation or on alternative projects, are scarce. The by far most important part of the costs of biodiversity conservation are the opportunity costs which result from the fact that, in order to conserve biodiversity, one cannot use land in alternative forms, for example for agriculture, developing rivers as water highways, etc.

7.3 The Economic Value of Biological Diversity

From what has been said in Section 7.2 above – namely (i) biodiversity satisfies human needs and (ii) biodiversity is scarce – it follows that biodiversity can be thought of as an economic good. Hence, one can attribute economic value to it. Conceptually and practically, the concept of economic value serves to capture the scarcity of biodiversity with respect to its potential to satisfy human needs. Valuation facilitates the aggregation of information in complex

[4]The economic definition of scarcity, based on (opportunity) costs, is one of *relative* scarcity. One may argue that biodiversity – like other natural resources – is not only scarce in a relative sense, but also in an absolute sense (Baumgärtner et al., in press). This goes beyond an economic analysis.

situations and, thus, is an important prerequisite for making rational decisions about the efficient allocation of resources (see Section 7.5 below).

Before I shall discuss the economic value of biodiversity in detail in Section 7.3.2 below, I shall first discuss the economic notion of value in general. This should help to understand the potential, but also the limits, of economic valuation of a natural resource, such as biodiversity.[5]

7.3.1 The Notion of Economic Value

When economists speak of a (material or immaterial) good's 'value', in most cases they mean an *instrumental* value. That is, the value of this good consists in it being a useful instrument in order to reach a certain goal. In contrast, one could ascribe an *intrinsic* value to a good.[6] That is, something could be valuable in itself, which is independent of it being an instrument in order to reach a certain goal.[7]

From the definition and the limitation of economics as the science which studies human behavior with regard to the satisfaction of human needs from scarce resources (Section 7.2), it is apparent that the satisfaction of human needs is the goal for which something should be instrumental, for it to have economic value. Thus, economic value is – by definition of economics – *anthropocentric*.

The methodological procedure, which economics follows in order to explain value, is that of so-called *methodological individualism*. In this approach, single individuals and the decisions and actions which result from their individual preferences and constraints, are taken as the elementary building blocks of explanation. In this perspective, the value of a good is ultimately determined by the interaction of the subjective valuations and actions of the many individuals in the economy. This means that economic value is determined by the subjective valuations of individuals in a society – and not, say, by the scientific

---

[5]The ethical and theoretical principles of economic valuation of environmental goods, services and damages are thoroughly discussed, for example, by Freeman (2003), Hanley and Spash (1993), Johansson (1987, 1999), Marggraf and Streb (1997).

[6]For example, Pirscher (1997) argues that biodiversity has an intrinsic value.

[7]The distinction between *instrumental* and *intrinsic* value correponds to a distinction already made by Immanuel Kant in his *Groundwork of the Metaphysics of Morals* (*Grundlegung zur Metaphysik der Sitten*, 1996[1785]: 84) between 'price' and 'dignity': 'In the kingdom of ends everything has either a price or a dignity ("Würde"). What has a price can be replaced by something else as its equivalent; what on the other hand is raised above all price and therefore admits of no equivalence has a dignity. What is related to general human inclinations and needs has a market price; that which, even without presupposing a need, conforms with a certain taste, that is, with a delight ("Wohlgefallen") in the mere purposeless ("zwecklosen") play of our mental powers, has a fancy price ("Affectionspreis"); but that which constitutes the condition under which alone something can be an end in itself has not merely a relative worth, that is, a price, but an inner worth, that is, dignity.'

judgment of experts.[8]

From this perspective it becomes obvious that economic value is not an inherent property of a commodity. Rather, it is attributed to a commodity by economic agents. What particular economic value is attributed to a commodity, hence, does not only depend on the objective (e.g. physical or ecological) properties of this thing, but also essentially depends on the whole socio-economic *context* in which valuation takes place. For example, when valuing natural resources such as clean drinking water, besides questions such as 'What is the utility of clean drinking water?' it is also important to consider questions such as: How much clean drinking water is there altogether? How is this amount distributed spatially and temporally? What are the institutions governing the access to the resource? What are the alternative uses besides use as drinking water, and what are the respective institutional constraints? Are there any alternatives to water in its different uses, and what are the respective conditions of provision?

The currently accepted paradigm of economics is neoclassical value theory.[9] According to this theory, value is a *marginal* concept: what is being valued are small changes in the state of the world (and not a certain state of the world), starting from the current state of the world. Thus, economic value is crucially determined by the current state of the world. This includes the current level of consumption of all different goods, current preferences, the current distribution of income and wealth, the current state of the natural environment, current production technology, and current expectations about the future – irrespective of whether we would think of this state of the world as being good or bad. Hence, the value of some good – in the neoclassical view – is the value of one additional unit of that good, given the amounts that we already consume of this good and all others. As a result, the value of some good is not constant, but changes with the amount already consumed of that good.

All these characteristic properties of the economic value concept, i.e. that economic value is

- instrumental,
- anthropocentric,
- individual-based and subjective,
- context dependent, and

[8]One problem which results from this individualistic approach, in particular when evaluating natural resources and environmental quality, is the aggregation of different subjective valuations to a social valuation (Seidl and Gowdy 1999: 106).

[9]The neoclassical value theory emerged in the so-called 'marginal revolution' around 1870, replacing the classical theory of value (see e.g. Blaug 1996 or Niehans 1990).

- marginal and state-dependent,

also apply to the economic value of biodiversity (Goulder and Kennedy 1997, Hampicke 1993, Nunes and van den Bergh 2001, Nunes et al. 2003, Seidl and Gowdy 1999, Weimann and Hoffmann 2003). While this focus allows economists to make clear and strong statements about the allocation of natural resources, its narrowness is a potential problem when linking with other academic disciplines in an encompassing discussion of biodiversity loss and conservation.

7.3.2 The Concept of Total Economic Value

In Section 7.2, I have listed a number of examples of how biodiversity satisfies human needs. Economists have tried to completely classify the different uses of the resource by the concept of *total economic value* (Pearce 1993, Pearce and Turner 1990: 129, Turner 1999b). This concept can be applied to the valuation of biodiversity (Watson et al. 1995a: 830ff., Geisendorf et al. 1998: 176ff., McNeely 1988: 14ff.). The total economic value of biodiversity, as an encompassing concept of the different human uses and motives for appreciation of biodiversity, may be classified into use and non-use values. Use values comprise all those value aspects which stem from actual or potential use of biodiversity. In contrast, non-use values are completely independent of any actual or potential use by the valuing individual (Krutilla 1967, Weisbrod 1964). They stem, for example, from the ethical, spiritual or religious desire to conserve biodiversity for the future, or for its own sake. On a more detailed level, use values may be classified into direct use value, indirect use value and option value. Non-use values may be classified into vicarious use value, bequest value and existence value.

Direct use value

Biodiversity has a direct use value insofar as different species and organisms, or parts thereof, directly satisfy human needs. On the one hand, this includes *consumptive use*, e.g. as food, fuel wood or medicinal plants, as well as *productive use*, e.g. as industrial resources, fuel or construction material (see the discussion in Section 7.2 above). On the other hand, this also includes *non-destructive use*, e.g. recreation, tourism, science and education (see the discussion in Section 7.2 above).

Indirect use value

Biodiversity has an indirect use value for humans insofar as biodiversity plays an important role in maintaining certain ecosystem services (Fromm 2000, Hueting et al. 1998) which, in turn, directly satisfy human needs or support economic processes that ultimately lead to the satisfaction of human needs. Examples (discussed in Section 7.2 above) include the support of biological productivity in agro-ecosystems, climate regulation, maintenance of soil fertility, control of water runoff, and cleansing of water and air.

Option value

Even if humans did not actually use biodiversity today, there is a value in the option of doing so tomorrow. This constitutes the option value of biodiversity. For example, the future may bring human diseases or agricultural pests which are still unknown today. Today's biodiversity would then have an option value insofar as the variety of existing plants may already contain a cure against the yet unknown disease, or a biological control of the yet unknown pest (Heal et al. 2004, Polasky and Solow 1995, Polasky et al. 1993, Rausser and Small 2000, Simpson et al. 1996, Swanson and Goeschl 2003). In this sense, the option value of biodiversity conservation corresponds to an insurance premium (Perrings 1995a, Weitzman 2000), which one is willing to pay today in order to reduce the potential loss should an adverse event – such as a human disease or an agricultural pest – occur in the future.[10]

While, strictly speaking, the conservation of anything has an option value, it is of particular importance in the case of biodiversity for two reasons. First, the loss of biodiversity is irreversible. Second, there is still large uncertainty about the different potential uses of biodiversity, e.g. as a storage of effective pharmaceutical substances or of desired genetic properties for crop varieties. Economists have stressed that under uncertainty it may be advantageous to postpone an irreversible decision, while gathering more information and learning (Arrow and Fisher 1974, Henry 1974).[11] In the case of biodiversity, such an option value clearly exists and may be considerable (Fisher and Haneman 1986, Weikard 2003).

Vicarious use value

The vicarious use value of biodiversity (Watson et al. 1995b: 13) is given by people's willingness to pay (or to forgo benefits) to ensure that other members of the present generation can enjoy the use value of biodiversity or specific components thereof. This is a form of altruism towards friends, relatives or strangers.

---

[10] The insurance value of biodiversity is subject to a detailed discussion in Chapters 9 and 10.

[11] This part of option value is often called 'quasi-option' value. It indicates the value of the additional information gained from postponing an irreversible decision and learning under uncertainty (Hanemann 1989).

Bequest value

The bequest value of biodiversity is given by people's willingness to pay (or to forgo benefits) to ensure that future generations can enjoy the use value of biodiversity or specific components thereof (Pommerehne 1987: 175ff.). This is a form of altruism towards future generations.

Existence value

The so-called existence value (Krutilla 1967: 781) of biodiversity is given by people's willingness to pay (or to forgo benefits) to ensure the continued existence of biodiversity or specific components thereof, irrespective of any actual or potential use by present or future generations of humans. This expresses an appreciation of biodiversity which is completely independent of any actual or potential, present or future use. It stems from a person's satisfaction from merely knowing that a particular species or ecosystem exists at all. It may be seen as a form of altruism towards non-human species or nature in general, and, in most cases, rests on ethical or religious motives. An indicator of the high importance of existence values may be the donations collected by nature conservation organizations for, say, the protection of the Siberian Tiger or the Panda Bear (Pearce and Turner 1990: 135).[12]

7.3.3 Methods for Identifying the Total Economic Value

At this point, one may summarize: The total economic value of biological diversity comprises very different components, corresponding to the very different human needs which are being satisfied by this natural resource. The different components of total economic value are, in principle, additive; but one needs to take care to not add mutually exclusive values (Moran and Pearce 1997: 2). For example, it would be a mistake to add the revenue from selling timber after clear-cutting a forest with the revenues of other (e.g. recreational) uses of the forest, because the latter are being destroyed by clear-cutting.

How to determine the total economic value? For goods which are being traded on markets one can (under certain conditions) take the market price as expressing the total economic value. For biological diversity, however, like with most natural resources, there is the problem that the resource is not, or only partially, being traded on markets. In order to determine its total economic

[12]Because the existence value is completely independent of any actual or potential human use, it does not seem to be an instrumental and, thus, economic value component. Indeed, the existence value is sometimes classified as an intrinsic value (e.g. by Watson et al. 1995b: 13, Pearce and Turner 1990: 130). However, while the existence value is independent of any actual or potential human use, it is not independent at all of the valuing economic agent (Pirscher 1997: 74). Knowing about the existence of a certain species is of utility for that economic agent. Thus, the existence value is an instrumental value, in so far as the existence of a certain species is instrumental for the utility for this economic agent.

value, or individual components thereof, one therefore needs to employ (direct or indirect) methods for non market valuation. These methods can, in principle, also be used to determine the total economic value of biodiversity (Watson et al. 1995a: 844–858).[13] Examples include the replacement cost method, the averting expenditure/avoiding costs method, the production function method, the hedonic pricing method, the travel cost method, or the contingent valuation method.[14]

7.4 Economic Causes of Biodiversity Loss

Extinction of species is not a new phenomenon. At all times, ever since life began on Earth, some species have become extinct and, at the same time, other species have originated. What is new today, however, is the high rate of species extinction, which is currently far above the long-term average rate known from fossil records. According to conservative estimates, the global rate of species extinction – averaged over all groups of species and ecosystems – currently varies between 50 and 100 times the natural rate (Watson et al. 1995b: 2). In tropical rain forests the extinction rate is considerably higher. It currently exceeds the natural rate of the past by a factor of somewhere between 1,000 to 10,000 (Watson et al. 1995b: 2). The large range of estimates indicates the considerable uncertainty about the exact number of extant species. The currently observed loss of biological diversity at all levels – genes, populations, species, ecosystems – is so dramatic that it may be considered the 'sixth mass extinction' (Wilson 1992: 32, Watson et al. 1995b: 22) in Earth's history.

The specific mechanisms through which the loss of populations, the extinction of species, and the impairment of ecological communities proceeds, are the following, listed in the order of their global importance (Watson et al. 1995b: 20):

1. loss, fragmentation, and degradation of habitats,
2. overuse of populations,
3. introduction of non-native species,
4. pollution of soil, water and air,

---

[13]Nunes and van den Bergh (2001) as well as Pearce and Moran (1994: 48) stress the considerable difficulties which occur when using these methods for the valuation of biological diversity.

[14]A detailed description of these methods would go beyond the scope of this chapter. For an introduction into these methods see, for example, Bateman et al. (2002), Freeman (2003), Hanley and Spash (1993), Pommerehne (1987) or Smith (1996).

5. climate change.

While in continental ecosystems the loss, fragmentation, and degradation of habitats is the most important mechanism, in oceanic ecosystems the overuse by fisheries and pollution are the most important factors. Coral reefs, which are a hot spot of biological diversity, are particularly affected by climate change. On islands, the introduction of non-native species and habitat loss are equally important.

The five specific mechanisms listed above are *proximate causes* of biodiversity loss. The underlying *primary causes* can be analyzed based on the concept of total economic value, which has been introduced in Section 7.3.2 above. Such an economic analysis is based on the identification of incentive structures which govern individual and social behavior in concrete situations. From such an economic perspective, the following four primary causes of biodiversity loss can be identified (Watson et al. 1995a: 830–832, Moran and Pearce 1997: 83–89):

1. population growth,
2. market failure,
3. governance failure,
4. fundamental ignorance.

These are discussed in detail in the following.

7.4.1 Population Growth

One cause of biodiversity loss, which seems most obvious, is the continuous population growth – though with decreasing rates – in developing countries and the continuous growth of the economy in the industrialized countries (Ehrlich and Holdren 1971, Holdren and Ehrlich 1974, Smith et al. 1995). Both developments imply an increasing demand for biological resources, and an ever increasing pressure on development of land as industrial space, for infrastructure (housing, highways, airports, etc.), or as agricultural land.

It seems inevitable that population growth and economic growth lead to a loss of biological diversity. The reason is the fundamental competition in land use: land can either be left in a natural state, thus serving as habitat for populations of wild species, or it can be developed for economic use, which means a loss of habitat for the originally living populations and, thus, their extinction. Since the land area on this planet is limited, population growth and economic growth necessarily mean – everything else being constant – that the economic use of land is being attributed a higher value due to increased demand, while its value as a natural habitat remains constant. As a result, more and more land is developed for economic use, which implies a continuous loss of biodiversity.

But this development is not as necessary as it may, at first, appear (Swanson 1995b). For, in the trade-off between the two fundamental alternatives – conservation of biological diversity versus economic development – there are many distortions, with the result that this trade-off is systematically biased in favor of economic development and against conservation of natural habitats. One may argue that the current loss of biodiversity is not primarily caused by population growth or economic growth, but rather by such distortions, which are now discussed in detail.

7.4.2 Market Failure

One standard result of economic theory is that the equilibrium on a competitive market, under certain conditions, is socially optimal in the sense that it is not possible to improve one individual's well-being without worsening some other individual's. One of the conditions under which this result holds is the absence of externalities. This means, all consequences of the transaction are mutually agreed upon by market participants and are being reflected in the market price. In contrast, if there are externalities, that is, if the market price does not reflect all consequences of a transaction, then markets may fail: the market price of a good does not reflect the total economic value of the good and the market equilibrium is not socially optimal. Externalities are ubiquitous in the allocation of biological resources.

Externalities due to incompletely specified property rights and missing markets

An externality arises if property rights for biological resources are only incompletely defined, or not defined at all (Lerch 1996, 1998). In the case of completely missing property rights or utilization rules, for populations of fish beyond the national coastal zones, for example, there is open access to a resource. It should be apparent that a resource which is useful and scarce but can be accessed without limit will, as a general rule, be overused. The case is similar for a resource which is being utilized by a number of individuals as a group, for example a village community, without any mandatory utilization rules. This has been described by Hardin (1968) as 'the tragedy of the commons'. Individual users of such a resource have an incentive to overuse the resource, because the benefits completely accrue to the individual while the problems stemming from overuse have to be carried by the whole group of users, and thus, only to a certain fraction by the individual user.

Market failure stemming from missing or incompletely specified property rights may be cured, in principle, by defining and enforcing property rights

(Swanson 1994). This is the logic behind the principle, expressed in the UN Convention on Biological Diversity (CBD 1992), that biological resources are the 'property' of the country in which these resources are located, and a similar proposal by the World Trade Organization (WTO) to introduce and enforce property rights in the form of patents ('intellectual property rights') on animal and plant genes (Sedjo and Simpson 1995, Swanson and Goeschl 2000b). The hope is that the loss of genetic diversity may be halted by attributing an adequate value to hitherto freely accessible, and thus undervalued resources, and to make the new resource owners manage these resources appropriately in their own interest. An example which illustrates the workings of this mechanism, is the agreement signed in 1991 between Costa Rica's National Institute for Biodiversity and the U.S. pharmaceutical company Merck Inc. (see Sedjo and Simpson 1995: 84ff., Lerch 1998: 292f.),[15] which has stimulated a host of similar agreements.

Character as public good

While biological resources have partly the character of a normal *private economic good* (as food or industrial resource, for example), in many other important respects they have the character of a *public good*. This means that (i) the use of the resource by one individual does not restrict or diminish the possibilities of use by another individual (*non-rivalry*), and (ii) no individual may be excluded from utilizing the resource (*non-excludability*). These two properties hold, in particular, for the important function of biodiversity in the provision of life supporting ecosystem services for humans, such as regulation of atmospheric composition or control of water circulation. The fact that one individual uses the constant oxygen fraction of the atmosphere for breathing does not restrict the possibility of other individuals to do the same. Furthermore, it is not possible to exclude individuals from using the atmosphere's oxygen for breathing.

The allocation of a public good on a competitive market is generally suboptimal (Varian 1992: Chapter 23), that is, there is market failure. The reason is that because of non-rivalry in consumption every single individual has an

---

[15] The Instituto Nacional de Biodidiversidad (INBio) of Costa Rica – a private nonprofit organization, which was founded following a recommendation of the government of Costa Rica – aims at conserving Costa Rica's biological wealth by promoting its intellectual and economic use. The agreement signed in 1991 and prolonged in 1994 and 1996 between INBio and the U.S. pharmaceutical company Merck Inc. states that Merck, at the beginning of each two-year term, makes a one time payment of 1 million U.S. dollars to INBio to support the Institute's work and the conservation of the natural rain forest, and obtains a certain number of plant specimens from the forest in exchange. In addition, Merck pays as a royalty to INBio a certain percentage of its sales from products which have been developed from these genetic resources. The contracting parties have stipulated non-disclosure of the level of these royalties. It is estimated that they are between 1% and 5%.

incentive to 'free ride', that is, to not reveal his or her true preferences for the consumption of the public good but to consume the amount of the good provided by other individuals without contributing to financing its provision. As a result, there is an under-provision with the public good on a competitive market compared to the social optimum. This particular form of market failure cannot be cured by defining private property rights in the public good because, due to the special character of these goods, in particular the property of non-excludability, the definition of property rights is, as a matter of principle, not possible.

Intragenerational spatial externalities: Global values vs. local markets

Many of the value components of biodiversity identified in Section 7.3.2 are global, for instance the vicarious use value, the bequest value, the existence value, but also the indirect use value which stems, for example, from the complex ecosystem of the Amazonian rain forest regulating weather and climate patterns on a global scale. This means, a large part of these values is appreciated by humans who do not live in the place where the resource is and, hence, do not take part in local decisions, as about land use in the Amazonian rain forest and its potential transformation into agricultural land. Put the other way round, the total economic value of biodiversity is not fully taken into account in local decisions about land use. In a local decision, the value of agricultural land use is compared with the value of land use as primary rain forest such as it is perceived by the local population. The latter is certainly much lower than the globally aggregated total economic value of primary rain forest. The externality consists in the fact that in a potential decision to clear-cut primary rain forest, the valuations of many of those who are affected by the transaction, namely the non-local users of the global public good, are not taken into account. Compared with the global total economic value, the value of biological diversity considered in the local decision is too low. As a result, too large a share of primary rain forest is transformed into agricultural land.

Intergenerational externalities: Present vs. future costs and benefits

A similar argument applies to the discrepancy of present and future costs and benefits of biodiversity. Today's markets only take into account today's decision makers' (expectations of) costs and benefits of transactions. Hence they neglect the part of total economic value which is due to future users who cannot take part in current decision processes.

Summary: Market failure

In the case of biodiversity a great number of different externalities act in the same direction. Today's market prices for essential components of biological diversity are considerably below their socially optimal value, which is given by the (globally and intertemporally aggregated) total economic value and comprises, besides the direct use value, also indirect use values and non-use values. In some cases, the (implicit) market prices are even zero. As a result, the private (opportunity) costs of, say, transforming primary rain forest into agricultural land, are far below the true (opportunity) costs which accrue to society at large. In turn, the private profits which can be made from transforming primary rain forest into agricultural land, are far above the profit for society at large. As a result of many externalities, unregulated markets lead to a rate of biodiversity loss which is too high compared with a social optimum (Swanson 1994).

7.4.3 Governance Failure

Many of the problems that fall under the heading 'market failure' could be cured, in principle, by adequate regulation of market processes. For example, one could propose that the global costs of loss of primary rain forest, which are not considered in local decisions, can be compensated for by a Pigouvian tax on tropical timber. The tax rate should be such that it covers the social costs of deforestation which are not currently included in market prices. The responsibility for such regulation is with the sovereign political bodies at all levels of organization – from the community level to the level of states and the international community of states. Failure to regulate in order to correct for market failure, is a form of *governance failure*.

Not only is governance failure widespread in current environmental problems, because regulative corrections of market failure are largely absent or not carried out appropriately, but some countries even reinforce market failure by policies which make market prices deviate even further from socially optimal prices. Examples include the subsidies for 'land cultivation' paid in Brazil for the clear-cutting of primary rainforest (Binswanger 1991) and the subsidies for offshore fisheries by the European Union.

An additional cause of biodiversity loss is the extreme inequality of income and wealth between the industrialized OECD countries of the North and the developing countries of the South, or more precisely, the poverty in rural areas in poor developing countries (Dasgupta 1993, 1995, Munasinghe 1992, Myers 1995, Swanson and Goeschl 2000a). The by far largest part of currently known biological diversity is found in the poorest countries of this world, namely in the equatorial regions of South America, Africa and Southeast Asia.[16] While

---

[16]Tropical rain forests host an estimated 50%, or even more, of all existing species on only 6% of the land area of the Earth (Myers 1995: 111). Tropical forests are currently destroyed

biodiversity protection in the OECD countries, such as by the establishment of nature protection areas, means an only moderate renunciation of (agricultural or industrial) economic use compared with total economic activity, for the rural population in the poorest countries of the world there is simply no decision problem 'nature protection versus economic use'. A renunciation of agricultural use, which is the sole source of income and food, would amount to starvation and, thus, represents no option at all. In so far as one considers it to be a task of responsible governance to establish international distributional justice, there is also a form of governance failure here.

7.4.4 Fundamental Ignorance

So far in this section, I have discussed causes of biodiversity loss as if the total economic value of biodiversity was perfectly known. In fact, however, it is not exactly known. Even if there should be scientific progress in revealing different components of this value, fundamental ignorance will remain because total economic value depends, inter alia, on the potential future use and on the indirect use of the resource. Particularly, with respect to these two potential uses of biodiversity a fundamental, 'irreducible' ignorance exists (Faber and Proops 1998: Chapter 7). It is still a largely open question, for example, what the exact role of biodiversity is for the stability of ecosystems and for the generation of different ecosystem services of human value (e.g. Holling et al. 1995, Hooper et al. 2005, Kinzig et al. 2002, Loreau et al. 2001, 2002b, Schulze and Mooney 1993, Tilman 1997a). Due to the high complexity of ecosystems one can safely assume that this ignorance cannot be reduced, even by intensive research, so much that more accurate predictions are possible about how human interference with biodiversity influences the functioning of ecosystems and the provision of ecosystem services. But this would be needed for a proper valuation of biodiversity.

This raises doubts about the relevance of the concept of total economic value as a basis for decision making, if individual components are subject to fundamental ignorance. Is this concept more than just a taxonomy? Indeed, acknowledging the fundamental ignorance about the exact quantity of biodiversity's total economic value could lead to the conclusion that society – rather than seeking the 'optimal' allocation of biodiversity – should follow a policy of 'safe minimum standards' (Ciriacy-Wantrup 1965), that is, set limits to habitat destruction so as to avoid the irreversible loss of critical biodiversity and ecosystem services.

---

at a higher rate than any other large-scale biome.

7.5 Relevance of Economic Valuation for the Protection of Biodiversity

As far as the pace and dimension of current biodiversity loss is concerned, the *Global Biodiversity Assessment* – a report from the United Nations Environment Programme – reaches the following conclusion: 'Because of the world-wide loss or conversion of habitats that has already taken place, tens of thousands of species are already committed to extinction. It is not possible to take preventive action to save all of them' (Watson et al. 1995b: 2). This conclusion contrasts markedly with naive ideas – which govern the thinking of many environmentalists and, to some extent, have influenced environmental legislation,[17] – that nature and species protection should aim at protecting *all* endangered species. But if it is not possible to protect *all* species which are at the brink of extinction today, the question is: what part of biodiversity should be protected? This covers two aspects which require trade-offs at different levels:

1. How important is the protection of one endangered species compared to another one?
2. How important is the protection of biological diversity compared with other societal goals?

These questions can be answered only unsatisfactorily, because any answer implies that a certain fraction of the currently existing biological diversity will become irreversibly extinct. But decisions have to be made, which means that these questions have to be answered – one way or the other. The economic value of biodiversity (cf. Section 7.3 above) offers a conceptual framework for discussing and answering these questions in a rational manner, based on the full set of available scientific knowledge.

7.5.1 Relative Importance of Different Biodiversity Protection Goals

So far, economic considerations only play a minor role in nature protection. The legislation on nature and species protection requires enforcement authorities to make recourse mainly to ecological and natural science concepts and evidence when classifying species as endangered – a decision which automatically entails efforts to protect these species. Economic considerations were not taken into account until recently.[18] As a result, for the listing of species as endangered, and thus worthy of being protected, it does not matter at all that different

---

[17]For example, this idea is explicated in the U.S. *Endangered Species Act* of 1973. See Brown and Shogren (1998) for an economic analysis of this legislation and its implementation.

[18]When U.S. Congress enacted the first version of the *Endangered Species Act* in 1973, they expressed explicitly that economic criteria are neither relevant for classifying species

species have different value for humans, and that the effective protection of different species comes at different costs. A species the protection of which is very expensive and which nevertheless has only moderate value, is treated on the same footing as a species with high economic value and relatively low costs of protection.

While the official rhetoric of nature and species protection starts from the premise that all endangered species are to be protected and, hence, does not include any explicit prioritization, time and budget constraints of the relevant enforcement authorities nevertheless require the setting of such priorities. Often, this is done only implicitly. Metrick and Weitzman (1996, 1998) studied, what criteria have actually determined the decisions of the U.S. *Office of Endangered Species* to classify a species as endangered, as well as public expenditures between 1989 and 1993 for the protection of individual species.[19] They found that the most important explanatory variables for listing a species as endangered were the degree of endangeredness, the distinctiveness of the species from other species (that is, its taxonomic uniqueness) and its size. Accordingly, mammals and fish ranged ahead of amphibians and reptiles. Spiders and insects, which make up the highest number of endangered species, are almost not represented on the list at all.

Metrick and Weitzman (1996, 1998) also found that, in contrast to listing a species as endangered, the expenditure for species protection measures correlates negatively (!) and significantly with the degree of the endangeredness of a species. There is a positive and significant correlation, however, with the body size of the animal. Expenditures on protection are also positively correlated with the status of the species as mammal or bird, and negatively correlated with the status as amphibian or reptile. Metrick and Weitzman (1998: 32) explain these results, in particular the surprising negative correlation of expenditures and endangeredness, by pointing to unobservable charismatic factors which are negatively correlated with the endangeredness, and which were not taken into account in their study. In this context, they speak of 'charismatic megafauna' – large and popular animals – which is obviously a decisive criterion for the willingness to spend money for the protection of the species. More than 50% of expenditures have been made for only 10 species (among which are the grizzly bear and the heraldic animal of the USA, the bald eagle). 95% of expenditures were in favor of vertebrate species. These numbers suggest that

---

as endangered, nor for setting up critical habitats (Brown and Shogren 1998: 4). The U.S. Supreme Court confirmed this view in 1978 in a leading decision (*Tennessee Valley Authority v. Hill*, 437 U.S. 187, 184 (1978)): 'the value of endangered species is incalculable' and 'it is clear from the Act's legislative history that Congress intended to halt and reverse the trend toward species extinction – whatever the cost'.

[19] Over these five years, a total of 914 million U.S. dollars has been spent on the protection of 229 vertebrate species (Metrick and Weitzman 1998: 28). The analysis of Metrick and Weitzman only studied expenditures which could be attributed to individual species.

an emotional identification with certain animals is actually more important for the protection decision than rational considerations based on scientific evidence and transparent criteria.

But if the goal should be to trade-off different alternatives and to set priorities based on scientific evidence and transparent criteria, the tool box of economics can be very helpful. Recall (from Section 7.2) that economics studies the optimal allocation of scarce resources from the point of view of society. Economic valuation of (ecological and economic) costs and benefits of different alternatives is one tool to assess the relative desiredness of different alternatives from the point of view of society. Economic valuation can therefore contribute to placing decisions on biodiversity protection on a rational and transparent basis (Dasgupta 2000, Weikard 1998b).

In particular, the total economic value of a species, which is broadly defined and covers, in principle, also the ecological functions of the species within an ecosystem, could help set up a rank ordering of species to be protected. Besides the value of the species, a prioritization should, of course, also be based on the costs of the protection measures for the species, and the increase of survival probability by this measure. Weitzman (1998) and Metrick and Weitzman (1996, 1998) have suggested – based on such reasoning and a formal economic analysis – the following simple criterion for calculating the rank of a species. Let V_i be the total economic value of species i, ΔP_i the increase in the species' survival probability by a protection measure, and C_i the measure's cost. Then a rank ordering in which different species are ranked according to the value of $R_i = V_i \cdot \Delta P_i / C_i$ is optimal from an economic point of view.[20] According to this criterion, the protection priority of the species is higher, the higher its total economic value, the more its survival probability can be increased by a protection measure, and the lower the costs of protection.

Of course, employing such a simple economic criterion for drawing up a rank ordering of species raises a number of questions. As already discussed above, to quantitatively determine the total economic value of a species is fraught with great difficulties. Also, one needs to be aware that this simple criterion is based on the assumption that species' survival probabilities are independent, which is clearly wrong when species interact in an ecosystem (Baumgärtner 2004c).[21] But the criterion allows one to enter a rational discussion about what information to use, and how, in order to prioritize among protection measures for different species. This is superior to the wide-spread current practice which

[20]Instead of total economic value V_i, Weitzman (1998: 1280) and Metrick and Weitzman (1998: 26) employ the sum $D_i + U_i$ of direct utility U_i and distinctiveness/uniqueness D_i of species i compared with other species. In so far as the latter gives rise to an indirect use value or an option/insurance value, both components are part of total economic value.

[21]How species interaction in an ecosystem influences optimal investment in species protection measures will be studied in detail in Chapter 11 below.

prioritizes only implicitly.

7.5.2 Relative Importance of Nature and Species Conservation Compared with Other Societal Goals

Concerning the question of the importance of the conservation of biodiversity compared with other societal goals, the essential economic idea is that biodiversity conservation requires the protection of natural habitats, that is, area of land. Land can be used for alternative purposes, for example as agricultural area, industrial space, or infrastructure, and is limited. It is therefore necessary to decide upon what share of land should be set aside for nature protection and what share is made available for economic development. The same goes for public or private financial resources. Put provocatively, the question is, how important is the conservation of biodiversity compared with social security, health care, education, etc.?

Even if one presupposes that the current loss of biodiversity is so dramatic that society is willing, based on the total economic value of biodiversity, to employ additional means for its protection, it is also obvious that, in principle, this trade-off can lead to the opposite result. That is, it could be that, at some point, society is not willing to sacrifice additional means for the protection of biodiversity because other purposes are considered to be more important. This means, an economic cost-benefit analysis can lead to the result that it is optimal to not protect a certain fraction of biodiversity, but let it become extinct, and employ the resources thus saved for achieving other societal goals.

7.5.3 The Design of Biodiversity Conservation Measures

Once a decision has been made upon (i) how important the protection of one endangered species is compared to another one, and (ii) how important the protection of biological diversity is compared with other societal goals, there is still the question of how to conserve biodiversity. This is the question of what instruments to apply, and how to design biodiversity conservation measures in order to reach a certain conservation goal. Any policy-relevant answer to this question must also rely on economics and economic valuation (Klauer 2001, Shogren et al. 1999).

With respect to in-situ conservation of species, the selection and design of reserve sites is the primary problem and, hence, has drawn most attention in the literature. Traditionally, the optimal selection and design of reserve sites is a domain of ecology and, in particular, its sub-field conservation biology (e.g. Margules et al. 1988, Soulé 1986). However, as Ando et al. (1998) and Polasky et al. (2001) have shown, cost savings of up to 80% could be achieved by integrating economic costs (i.e. land prices) into previously ecologically based selection algorithms for reserve sites. This demonstrates that economics and economic valuation are important for the design of cost effective and efficient measures of nature conservation. As a result, economists have turned to this issue and made a number of important contributions (e.g. Costello and Polasky 2004, Johst et al. 2002, Polasky et al. 1993, Solow et al. 1993, Weitzman 1993, Metrick and Weitzman 1996, 1998, Wu and Boggess 1999).

7.6 Conclusion: What can Economics Contribute to Biodiversity Conservation?

Economics is – like any scientific discipline – limited by its research interest and its methodology. It views biodiversity as a resource for the satisfaction of various human needs, and studies its efficient allocation for that sake (see Section 7.2). This also defines the economic value of biodiversity. In particular, the economic value notion is instrumental, anthropocentric, individual-based/subjective, context dependent, marginal and state-dependent (see Section 7.3). In spite of these limitations, economics can make important contributions to the study of biodiversity loss and conservation (Brown and Shogren 1998: 15–19).

First, the economic perspective gives a detailed and good understanding of the specific mechanisms and fundamental causes of the current dramatic loss of biodiversity (see Section 7.4). Besides human population growth, it is mainly different forms of market failure which make market prices on today's markets for biological resources deviate from their optimal levels. This results in a suboptimally high rate of biodiversity loss. Governance failure in regulating the access to, and use of, biological resources in a fair, efficient and sustainable manner is an equally important reason for this development. Another major reason is the fundamental ignorance about potential future uses of biological diversity and about the role that biodiversity plays for the functioning of ecosystems.

Second, the question of what species are threatened by extinction or will be in the near future, is not only an ecological question but also an economic question. For, besides ecological constraints, economic developments influence the extinction probability of a species as well. This extinction probability is higher for species which rival with economic development (e.g. highway construction leading to habitat fragmentation). It is lower for species which are under intensive protective care. Since the decision between economic development and nature protection is crucially determined by economic considerations, so is the probability of extinction of different species. Endangeredness, thus, is not a purely ecological quality, which could be determined solely from natural science research, but is crucially determined by economic factors as well.

Third, since not all species that are threatened by extinction today can be saved (Watson et al. 1995b: 2), the question arises 'What species and populations should be protected, and to what extent?' Economics can provide a methodological framework to discuss this question rationally. It can thereby help base decisions about biodiversity protection on scientific evidence and transparent criteria (see Section 7.5). This framework essentially builds on valuation and the comparison of alternative options based on their respective (ecological and economic) costs and benefits. Economics therefore allows one to prioritize among different protection goals.

Fourth, economics is indispensable when a given protection goal is to be achieved in a cost effective manner. This means, after a certain protection goal has been set (for example, the conservation of an endangered population in a certain region), one chooses among all protection measures that are suitable to actually achieve the goal (for example, displacing the highway, building bridges over the highway, establishment of a compensation habitat in a different location, etc.) the one with minimal costs.

As far as the first and the last points are concerned, economics is in its traditional domain and can make powerful contributions, even to the solution of a problem such as biodiversity loss, which is primarily defined in biological terms. As far as the second and third points are concerned, economic research is still in its infancy. While there are already a number of promising contributions, it has also become obvious that – in order to develop meaningful analyses and relevant policy recommendations – one needs an interdisciplinary cooperation between ecologists and economists, leading to fully integrated ecological-economic analyses of biodiversity loss and conservation (Wätzold et al. 2005).

8. Ecological and Economic Measures of Biodiversity

8.1 Introduction

For analyses of how biodiversity contributes to ecosystem functioning, how it enhances human well-being, and how these services are currently being lost, a quantitative measurement of biodiversity is crucial. Ecologists, for that sake, have traditionally employed different concepts such as species richness, Shannon-Wiener-entropy, Simpson's index, or the Berger-Parker-index (e.g. Magurran 2004, Pielou 1975, Purvis and Hector 2000). Recently, economists have added to that list measures of biodiversity that are based on pairwise dissimilarity between species (Solow et al. 1993, Weikard 1998a, 1999, 2002, Weitzman 1992, 1993, 1998) or, more generally, weighted attributes of species (Nehring and Puppe 2002, 2004).

The full information about the biodiversity of an ecological system is only available in the full description of the system in terms of the number of different entities (i.e. genes, species, or ecological functions), their abundances and characteristic features. Such a full description comes in different and complex statistical distributions. For the purpose of comparing two systems, or describing the system's evolution over time, both of which is essential for policy guidance, it seems therefore necessary to condense this information into easy-to-calculate and easy-to-interpret numbers, although that certainly means a loss of information. Most often, all the relevant information about the diversity of a system is condensed into a single real number, commonly called a 'measure of diversity' or 'diversity index'. As there are virtually infinitely many ways of calculating such a diversity index from the complex and multifarious information about the system under study, it is crucial to be aware of which aspects of information are being stressed in calculating the index and which aspects are being downplayed, or even neglected altogether. Not surprisingly then, the *purpose* for which a particular index is calculated and used is crucial for understanding how it is prepared.

In this chapter, I give a conceptual comparison of the two broad classes of biodiversity measures currently used, the ecological ones and the economic ones. It will turn out that there are systematic differences between these two classes, concerning exactly what information is being used to construct the index. For example, while ecological indices consider abundances, the economic ones deliberately do not take abundances into account. In order to explain these differences, I will argue that the two types of measures aim at characterizing two very different aspects of the ecosystem. One important observation is that the rationale for, and basic conceptualization of, the economic measures of biodiversity stems from the economic idea of product diversity and is intimately related to the idea of choice between different products which can, in principle, be produced in any given number. These measures characterize an abstract commodity/species space, rather than a real allocation of commodities/species. This raises a number of questions about the applicability of these concepts in ecology. I will conclude that the measurement of biodiversity requires prior value judgments as to what purpose biodiversity serves in ecological-economic systems.

The chapter is organized as follows. Section 8.2 introduces a formal and abstract description of ecosystems. This framework allows the rigorous definition and comparison of different biodiversity indices later on. Section 8.3 then addresses the question of how to quantitatively measure the biodiversity of an ecosystem by surveying different ecological and economic measures of diversity. Section 8.4 compares the different measures at the conceptual level, identifies essential differences between them, and critically discusses these differences. Section 8.5 concludes.

8.2 Species and Ecosystems

'Biodiversity' has been defined as 'the variability among living organisms from all sources [...] and the ecological complexes of which they are part' (CBD 1992), which encompasses a wide spectrum of biotic scales, from genetic variation within species to biome distribution on the planet (Gaston 1996, Groombridge 1992, Purvis and Hector 2000, Wilson 1992). In this chapter, I shall only be concerned with the level of species, as this is the level of organization which is currently being given most attention in the discussion of biodiversity conservation policies.[1] That is, biodiversity is here considered in the sense of species diversity.

In order to describe the species diversity of an ecosystem, and to compare two systems in terms of their diversity, one can consider different structural

[1]Ceballos and Ehrlich (2002) have pointed out that the loss of populations is a more accurate indicator for the loss of 'biological capital' than the extinction of species.

characteristics of the system(s) under study:

- the *number* of different species in the system,
- the characteristic *features* of the different species, e.g. their functional traits, and
- the relative *abundances* with which individuals are distributed over different species.

Intuitively, it seems plausible to say that a system is more diverse than another one if it comprises a higher number of different species, if the species in the system are more dissimilar from each other, and if individuals are more evenly distributed over the different species. A simple example can illustrate this idea

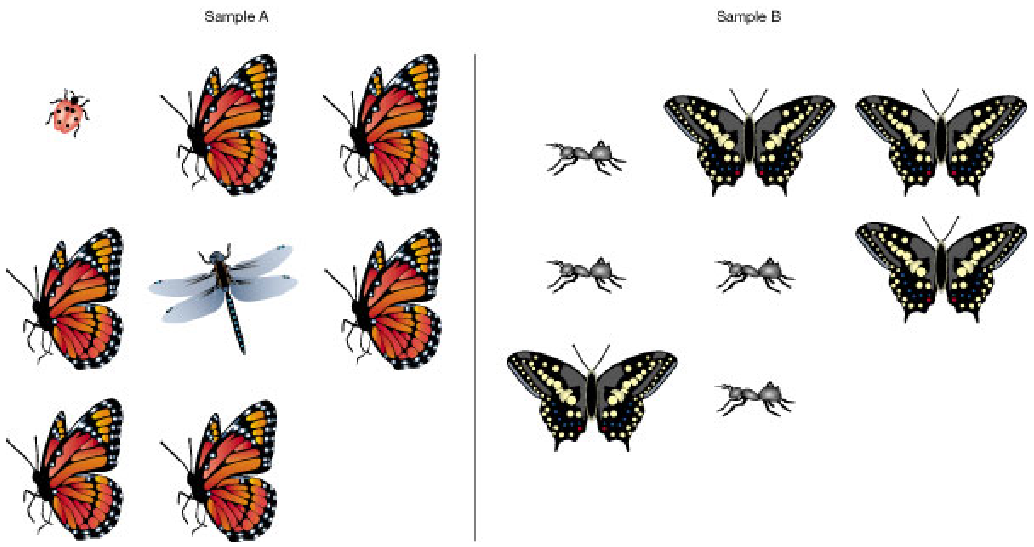

Figure 8.1 Two samples of species, which may be compared in terms of their diversity based on different criteria: species number, species abundances and species features. Figure taken from Purvis and Hector (2000: 212).

(Figure 8.1). Consider two systems, A and B, which both consist of eight individuals of insect species: system A comprises six monarch butterflies, one dragonfly and one ladybug; system B comprises four swallowtail butterflies and four ants. Obviously, according to the first criterion (species number), system A has a higher diversity (three different species) than system B (two different species). But according to the third criterion (evenness of relative abundance) one may as well say that system B has a higher diversity than system A, because there is less chance in system B that two randomly chosen individuals will be of the same species. And as far as the second criterion goes (characteristic species features), one would have to start by saying what the characteristic species features actually are, which can then be used to assess the aggregate dissimilarity of both systems.

Before discussing these ideas in detail, let me first introduce a formal and abstract description of the ecosystem whose species diversity is of interest. Let n be the total *number* of different species existent in the system and let $S = \{s_1, \ldots, s_n\}$ be the set of these species. Each s_i (with $i = 1, \ldots, n$) represents one distinct species. In the example illustrated in Figure 8.1, S^A = {monarch butterfly, ladybug, dragonfly } and S^B = {swallowtail butterfly, ant}. In the following, $n \geq 2$ is always assumed.

Let m be the total number of different relevant features, according to which one can distinguish between species, and let $F = \{f_1, \ldots, f_m\}$ be the list of these *features*. Each f_j (with $j = 1, \ldots, m$) represents one distinct feature. For example, possible features could include the following:

- being a mammal/bird/fish/...,
- being a herbivore/carnivore/omnivore,
- unit biomass consumption/production,
- being a 'cute little animal',
- etc.

Then one can characterize each species s_i (with $i = 1, \ldots, n$) in terms of all features f_j (with $j = 1, \ldots, m$). Let x_{ij} be the description of species s_i in terms of feature f_j, so that $x = \{x_{ij}\}_{i=1,\ldots,n;\ j=1,\ldots,m}$ is the complete characterization of all species in terms of all relevant features.

The *abundance* of different species in the ecosystem is described by the distribution of absolute abundances of individuals over different species. Let a_i be the *absolute abundance* of individuals of species s_i (with $i = 1, \ldots, n$). If the system under study contains only species of the same or a very similar kind, the abundance of a species in an ecosystem may be measured by simply counting the number of individuals of that species. In the example illustrated in Figure 8.1, $a^A = (6, 1, 1)$ and $a^B = (4, 4)$, i.e. in system A there are six individuals of species 1 (monarch butterfly), one individual of species 2 (ladybug), one individual of species 3 (dragonfly), and in system B there are four individuals of species 1 (swallowtail butterfly) and four individuals of species 2 (ant). However, if the system comprises species which are very different in size, e.g. deer, birds, butterflies, ants and protozoa, it makes very little sense to measure their respective absolute abundances by just counting individuals (Begon et al. 1986: 594). Their enormous disparity in size would make a simple count of individuals very misleading. In that case, the absolute abundance of a species may be measured by the total biomass stored in all individuals of that species.

Typically, it is more interesting to consider not the *absolute* abundance a_i of species s_i, but its *relative* abundance in relation to all the other species. The relative abundance of species s_i is given by $p_i = a_i / \sum_{i=1}^n a_i$. Let $p = (p_1, \ldots, p_n)$ be the vector of relative abundances. By construction of p_i, one has $\sum_{i=1}^n p_i = 1$ and $0 \leq p_i \leq 1$, where $p_i = 0$ means that species i is absent from the system and $p_i = 1$ (implying $p_j = 0$ for all $j \neq i$) means that species i is the only species in the system. In the example illustrated in Figure 8.1, $p^A = (0.75, 0.125, 0.125)$ and $p^B = (0.5, 0.5)$. If species abundances are measured by counting individuals of that species, the relative abundance p_i indicates the probability of obtaining an individual of species s_i in a random draw from all individuals in the system. When abundances are measured in biomass, the relative abundance p_i indicates the relative share of the ecosystem's biomass stored in individuals of species s_i. Without loss of generality, assume that $p_1 \geq \ldots \geq p_n$, i.e. species are numbered in the sequence of decreasing relative abundance, such that s_1 denotes the most common species in the system whereas s_n denotes the rarest species.

Altogether, the formal description of an actual or potential ecosystem state Ω comprises the specification of n, S, m, F, p and x, which completely describes the composition of the ecosystem from different species as well as all species in terms of their characteristic features. In the following, a *biodiversity measure* of the ecosystem Ω means a mapping D of all these data on a real number:

$$D : \Omega \to \mathbb{R} \quad \text{with} \quad \Omega = \{n, S, m, F, p, x\} \,. \tag{8.1}$$

That is, I consider only biodiversity measures which characterize the species diversity of an ecosystem by a single number ('biodiversity index').[2] The various measures differ in what information about the ecosystem state Ω they take into account and how they aggregate this information to an index.

8.3 Different Measures of Biodiversity

8.3.1 Species Richness

The simplest measure of biodiversity of an ecosystem Ω is just the total number n of different species found in that system. This is often referred to as *species richness* (following McIntosh 1967):

$$D^R(\Omega) = n. \tag{8.2}$$

[2]Note that the focus on biodiversity *indices* constitutes a considerable reduction in generality and has a significant economic bias. The desire to characterize a set of objects by a single number – instead of, say, by the distribution of properties or abundances – can be vindicated by the aim of establishing a *rank ordering* among different sets, which is necessary in order to *choose* the best – in the sense of: most diverse – set (e.g. Weitzman 1992, 1998).

Species richness is widely used in ecology as a measure of species diversity. One example is the long-standing and recently revitalized diversity-stability debate, i.e. the question whether more diverse ecosystems are more stable and productive than less diverse systems (Elton 1958, Odum 1953, Hooper et al. 2002, Loreau et al. 2001, 2002, MacArthur 1955, May 1972, 1974, McCann 2000).[3] Another example are the so-called species-area relationships,[4] which are important for the present biodiversity conservation debate because they are virtually the only tool to estimate the number of species that go extinct due to large-scale habitat destruction (Gaston 2000, Kinzig and Harte 2000, MacArthur and Wilson 1967, May et al. 1995, Rosenzweig 1995, Whitmore and Sayer 1992). Species richness is also the biodiversity indicator implicitly used in the public discussion, which often reduces biodiversity loss to species extinction.

In the species richness index (8.2), all species that exist in an ecosystem count equally. However, one might argue that not all species should contribute equally to an index of species diversity. Two different strands have evolved in the literature both of which develop indices in which different species are given different weight. The first strand, which has evolved mainly in ecology, weighs different species according to their relative abundance in the system. This is vindicated by the observation that the functional role of species may vary with their abundance in the system. These biodiversity indices are discussed in Section 8.3.2 below. The other strand, which has been contributed to the discussion of biodiversity mainly by economists, stresses that different species should be given different weight in the index due to the characteristic features they possess. These biodiversity indices are discussed in Section 8.3.3 below.

8.3.2 Indices Based on Relative Abundances

Ecologists have tackled the problem of incorporating the functional role of species in a measure of species diversity by formulating diversity indices in which the contribution of each species is weighted by its relative abundance in the ecosystem (Begon et al. 1998, Magurran 1988, Pielou 1975, Purvis and Hector 2000, Ricklefs and Miller 2000). Intuitively, rare species should contribute less than common species to the biodiversity – in the sense of 'effective' species richness – of an ecosystem.[5] A general measure for the effective number ν of species, which uses the information about pure species number n and the

---

[3]The diversity-stability relationship is tantamount for the insurance value of biodiversity (Chapter 9). See the detailed discussion on this relationship in Section 9.2.

[4]The well established species-area-relationships state that species richness n increases with the area l of land as $n \sim l^z$, where z (with $0 < z < 1$) is a characteristic constant for the type of ecosystem.

[5]Rao (1982) equates species richness and distribution of relative abundances with community size and shape respectively.

distribution of relative abundances $p = (p_1, \ldots, p_n)$ to build on this intuition, is the following (Hill 1973):[6]

$$\nu_\alpha(n,p) = \left(\sum_{i=1}^{n} p_i^\alpha \right)^{1/(1-\alpha)} \qquad \text{with } \alpha \geq 0 \ . \tag{8.3}$$

This measure has a number of desired properties, which have made it the foundation for various biodiversity indices in ecology:

1. The measure (8.3) – more exactly: its logarithm $H_\alpha = \log \nu_\alpha$ – is well known from information theory where it has been introduced by Rényi (1961) as a generalized entropy ('entropy of order α of the probability distribution p'). Its properties are well studied and understood (Aczél and Daróczy 1975).

2. The maximal value of $\nu_\alpha(n,p)$ increases with the number n of different species.

3. For given n, the measure ν_α takes on values between 1 and n, depending on p. Technically, it is an inverse generalized mean relative abundance: ν_α gives the equivalent number of equally abundant (hypothetical) species that would reproduce the entropy value H_α of the actual system of n species with unevenly distributed relative abundances p (Whittaker 1972). Thus, ν_α can be interpreted as an effective species number in a system of n unevenly distributed species.

4. For given n, the measure $\nu_\alpha(n,p)$ assumes its maximal value – that is, pure species richness n – when all species have equal relative abundance, i.e. $p_i = 1/n$ for all $i = 1, \ldots, n$. In this case of an absolutely even distribution of relative abundances, the effective number $\nu_\alpha(n,p)$ simply equals the total number n of different species in the system.

   The measure $\nu_\alpha(n,p)$ decreases with increasing unevenness of the distribution of relative abundances p. This means, dominance of a few species, or, more generally, an uneven distribution of relative abundances, brings down the index of effective species number $\nu_\alpha(n,p)$. The index assumes its minimal value when a system is dominated by one single species, with

[6]Neither Rényi (1961) nor Hill (1973), who introduced this measure to information theory and to ecology respectively, restrict the range of α to non-negative real numbers. Indeed, Equation (8.3) is well defined for all $-\infty \leq \alpha \leq +\infty$. However, for $\alpha < 0$, $\nu_\alpha(n,p)$ yields values greater than n, which means that rarer species are given greater weight than more common species in the measure of effective species number. This contradicts the intuition that the effective species number should be smaller than the pure number, depending on heterogeneity. Therefore, when it comes to measuring species diversity it seems to be reasonable to constrain the parameter α to non-negative values.

all other species having negligible relative abundances, i.e. $p_i \approx 0$ for all $i = 1, \ldots, n$ except for $i = i^*$, where i^* denotes the dominant species, and $p_{i^*} \approx 1$. In this case, $\nu_\alpha(n, p) \approx 1$, which means that the effective number of different species is approximately one.

5. The parameter $\alpha \geq 0$ weighs the influence of evenness of the distribution of relative abundances p against the influence of pure species number n on the effective species number ν. For $\alpha = 0$, the evenness of the distribution of relative abundances p is completely irrelevant, and the effective species number ν is simply given by the pure species number n. The larger α, the higher is the weight of the evenness in the calculation of the effective species number (8.3). For $\alpha = +\infty$, the pure species number n is completely irrelevant, and the effective species number ν is exclusively determined by how (un)evenly the relative abundances of species are distributed.

6. For different values of the parameter α one can recover from expression (8.3) different species diversity indices that are well-established in ecology (Hill 1973). They can thus be considered as special cases of the general measure (8.3):

Species richness index

With $\alpha = 0$ one obtains the *species richness index* already discussed in Section 8.3.1 above:

$$D^R(\Omega) = \nu_0(n, p) = n \ . \tag{8.4}$$

That is, to zeroth order effective species number is just pure species richness.

Shannon-Wiener index

With $\alpha = 1$ one obtains[7] the *Shannon-Wiener-index*

$$D^{SW}(\Omega) = \nu_1(n, p) = \exp H \quad \text{with} \quad H = -\sum_{i=1}^{n} p_i \log p_i \ , \tag{8.5}$$

where H is well known from statistics and information theory as the Shannon-Wiener expression for entropy (Shannon 1948, Wiener 1948).[8]

[7]This is not immediately obvious. See Hill (1973: Appendix) for a proof.

[8]The expression H (Equation 8.5) has been proposed independently by Claude Shannon (1948) and Norbert Wiener (1948). It is sometimes referred to as Shannon-Weaver-entropy because it has been popularized by Shannon and Weaver (1949). In information theory the base of the logarithm is usually taken to be 2, consistent with an interpretation in terms of 'bits'. In ecology the tendency is to employ natural logarithm's, i.e. a base of e, although some use a base of 10. There is, of course, no natural reason to prefer one base over the

Shannon-Wiener entropy, and the index built from it, does not have a straightforward, let alone ecologically meaningful, interpretation as Simpson's index has (see below). Being a logarithmic measure, it is also more difficult to calculate than Simpson's index. Nevertheless, it is a popular measure of heterogeneity and effective species number. This is especially due to the logic of its development within statistical physics (Balian 1991) and information theory (Krippendorff 1986), and its formal elegance and consistency. For example, of all the measures defined by the general expression (8.3) for $0 \leq \alpha \leq +\infty$, only Shannon-Wiener entropy ($\alpha = 1$) allows consistent aggregation of heterogeneity over different hierarchical levels of a system: upper level Shannon-Wiener entropy of a system of individuals clustered in lower level subsystems can be additively decomposed to show the contributions from heterogeneity within and between lower level subsystems.

Simpson's index

With $\alpha = 2$ one obtains *Simpson's index* (Simpson 1949):

$$D^S(\Omega) = \nu_2(n,p) = 1/\sum_{i=1}^{n} p_i^2 \ . \tag{8.6}$$

Simpson's index has been, and still is, fairly popular among ecologists. The reasons include the ease of calculating the index, the bounded properties of the expression $\sum p_i^2$, and – not the least – the ecological meaningfulness of its interpretation: $\sum p_i^2$ is the probability that any two individuals drawn at random from an infinitely large ecosystem belong to different species.[9] The inverse of this expression is taken to form the biodiversity index, so that D^S increases with the evenness of the distribution of relative abundances. This makes sense as an index of effective species number when viewing ecosystems as functional relationships, e.g. based on predator-prey-relations, parasite-host-relations, etc.

Berger-Parker index

With $\alpha = +\infty$ one obtains the *Berger-Parker-index* (Berger and Parker 1970, May 1975) as

$$D^{BP}(\Omega) = \nu_{+\infty}(n,p) = 1/p_1 \ , \tag{8.7}$$

---

other, but care should be taken when comparing results from different studies in terms of H, which might have been obtained using different bases. Yet, the choice of a particular base does not have any influence on ν (as long as one chooses the same base for the logarithm and the exponential function in Equation 8.5.)

[9] The appropriate formula for a finite community is $\sum[n_i(n_i-1)/(N(N-1)]$, where n_i is the number of individuals in the ith species and $N = \sum_{i=1}^{n} n_i$ is the total number of individuals.

that is, the inverse relative abundance of the most common species in the system. It can be interpreted as an effective species number in the sense that $1/p_1$ gives the equivalent number of equally abundant (hypothetical) species with the same relative abundance as the most abundant species in the system. Obviously, the Berger-Parker-index only takes into account the relative dominance of the most common species in the system, neglecting all other species.

7. One of the properties of the biodiversity measure (8.3) is that for given n and p the value of $\nu_\alpha(n, p)$ decreases with α. As the most widely used diversity indices can all be expressed as special cases of Equation (8.3) for different values of a, it becomes evident that the results for the effective species number in a given system yielded by these indices are related in the following way:

$$n = D^R \geq D^{SW} \geq D^S \geq D^{BP} > 1 \ . \tag{8.8}$$

Illustration

Table 8.1 illustrates the working of the various indices in comparison for different hypothetical communities. The first observation is that for all communities,

species s_i	relative abundance p_i in community					
	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6
s_1	0.25	0.20	0.24	0.249	0.50	0.50
s_2	0.25	0.20	0.24	0.249	0.30	0.30
s_3	0.25	0.20	0.24	0.249	0.10	0.10
s_4	0.25	0.20	0.24	0.249	0.07	0.07
s_5	-	0.20	0.04	0.004	0.03	0.01
s_6	-	-	-	-	-	0.01
s_7	-	-	-	-	-	0.01
n $(\alpha = 0)$	4	5	5	5	5	7
D^{SW} $(\alpha = 1)$	4.00	5.00	4.48	4.08	3.42	3.53
D^S $(\alpha = 2)$	4.00	5.00	4.31	4.03	2.81	2.82
$D^{BP}(\alpha = +\infty)$	4.00	5.00	4.17	4.02	2.00	2.00

Table 8.1 Comparison of different diversity indices for hypothetical communities Ω_j ($j = 1, \ldots, 6$) of four, five or seven different species with relative abundances p_i. The parameter α pertains to the general definition (8.3), n is the species richness of the respective community (Equation 8.4), D^{SW} is the Shannon-Wiener index (Equation 8.5), D^S is Simpson's index (Equation 8.6), D^{BP} is the Berger-Parker index (Equation 8.7).

$\alpha = 0$ yields species richness n as the effective species number of that community. Second, the Berger-Parker index, as the limit case of $\alpha = +\infty$, gives the number of equally abundant (hypothetical) species with the same relative abundance as the most abundant species in the community, $1/p_1$. If, for example, the most common species has a relative abundance of $p_1 = 0.5$, with the other species in that community having smaller relative abundances, then the effective number of species in that community would be $D^{BP} = 1/0.5 = 2$, irrespective of the number and relative abundances of the other species (Table 8.1, columns 6 and 7). Third, all indices – i.e. all values of α – yield species richness n as the effective species number if the community consists of absolutely evenly distributed species, for example in communities Ω_1 and Ω_2 (Table 8.1, columns 2 and 3). In this case, in index value is the higher the higher n.

Fourth, for a given number of species, e.g. $n = 5$, all indices assume their maximal value – species richness n – if species are absolutely evenly distributed, for example in community Ω_2 as compared to Ω_3, Ω_4 and Ω_5 (Table 8.1, columns 3 through 6). Conversely, the value of the index decreases if species are distributed more unevenly. The higher α, the stronger the index value decreases with unevenness. For example, comparing the five-species-communities Ω_2 and Ω_5 (Table 8.1, columns 3 and 6) shows that the index value drops from 5 to 2 for $\alpha = +\infty$, while it only drops to 3.42 for $\alpha = 1$ and remains at the level of 5 for $\alpha = 0$. Communities Ω_3 and Ω_4 (Table 8.1, columns 4 and 5) illustrate that with $n - 1$ species of equal relative abundance and one species, s_5, with much lower relative abundance, the Simpson, Shannon-Wiener and Berger-Parker indices will be only slightly greater than $n - 1$. The smaller p_5, the closer they approach $n - 1$.

Sixth, a comparison of communities Ω_2 and Ω_6 (Table 8.1, columns 3 and 7) shows that the effective species number as measured by ν_α can actually decrease although species richness, n, increases between two communities. This is due to the increase in unevenness outweighing the increase in species richness.[10]

Seventh, the higher α, the more weight an index puts on the more abundant species in the community while being less sensitive to differences in small relative abundances and in total species richness, as can be seen from comparing communities Ω_5 and Ω_6 (Table 8.1, columns 6 and 7). These two communities only differ in the number and composition of very rare species. The Berger-Parker index ($\alpha = +\infty$), which takes into account only the most abundant species, is completely insensitive to this difference. Even Simpson's index ($\alpha = 2$) is hardly sensitive to this difference. The Shannon-Wiener index ($\alpha = 1$) is more sensitive to differences in small relative abundances than Simpson's index,[11] but only species richness ($\alpha = 0$) fully takes into account the higher number of very rare species in community Ω_5 compared to Ω_6.

---

[10] May (1975) has shown that for $n > 10$ the underlying species abundance distribution makes a crucial difference for how, and even whether at all, D^S increases with n.

[11] On the other hand, it is less sensitive to small differences in large relative abundances, whereas Simpson's index responds more substantially to these differences.

Coming back to the simple example of comparing the two communities described in Section 8.2 (Figure 8.1), the different biodiversity indices discussed in this section yield the results shown in Table 8.2. As one can see from the

species s_i	relative abundance p_i in system A	system B
s_1	0.75	0.5
s_2	0.125	0.5
s_3	0.125	—
n $(\alpha = 0)$	3.00	2.00
D^{SW} $(\alpha = 1)$	2.09	2.00
D^{S} $(\alpha = 2)$	1.68	2.00
$D^{BP}(\alpha = +\infty)$	1.25	2.00

Table 8.2 Comparison of different diversity indices for the two systems described in Section 8.2 (Figure 8.1). The parameter α pertains to the general definition (8.3), n is the species richness of the respective community (Equation 8.4), D^{SW} is the Shannon-Wiener index (Equation 8.5), D^{S} is Simpson's index (Equation 8.6), D^{BP} is the Berger-Parker index (Equation 8.7).

table, which system is rated to be 'more diverse' depends on the parameter α, i.e. on how one weighs pure species richness against evenness of relative abundances: for small values of α sample A is found to be more diverse, while system B turns out to be more diverse for large values of α.

8.3.3 Indices Based on Characteristic Features

The biodiversity indices discussed in Section 8.3.2 all take the species richness of an ecosystem, properly adjusted by the distribution of relative abundances so that rare species are given less weight than common species, to be a measure of diversity. According to these indices, systems with more, and more evenly distributed, species are found to have a higher biodiversity than systems with less, and less evenly distributed, species. This procedure has been criticized for not taking into account the (dis)similarity between species. For example, a system with 100 individuals of some plant species, 80 individuals of a different plant species, and 50 individuals of yet another plant species will be found to have exactly the same biodiversity, according to these indices, than a system with 100 individuals of some plant species, 80 individuals of a mammal species, and 50 individuals of some insect species. Yet, intuitively one would say that the latter has a higher biodiversity. This intuition is based on the (dis)similarity between the various species.[12]

[12] The richness-and-abundance based indices discussed in Section 8.3.2 implicitly assume that all species are pairwise equally (dis)similar.

In order to account for the (dis)similarity of species when measuring biodiversity, one needs a formal representation of the characteristic features of species. Based on these characteristic features, the (dis)similarity of species can be measured and taken into account when constructing a biodiversity index. Two different approaches exist so far. One has been initiated by ecologists (May 1990, Erwin 1991, Vane-Wright et al. 1991, Crozier 1992) and put on a rigorous axiomatic basis, enhanced and popularized by Weitzman (1992, 1993, 1998). I shall therefore call it the Weitzman-approach.[13] It builds on the concept of a distance function to measure the pairwise dissimilarity between species. The diversity of a set of species, in this approach, is then taken to be an aggregate measure of the dissimilarity between species. This approach is most appealing when applied to phylogenetic diversity. The other approach, developed by Nehring and Puppe (2002, 2004), generalizes the Weitzman-approach. It builds directly on the characteristic features of species and their relative weights. Both approaches are now discussed in detail.

Weitzman index

Weitzman (1992) defines a diversity measure, $D(S)$, of a set S of species based on the fundamental idea that the diversity of a set of species should be an aggregate measure of the pairwise dissimilarity between species. The dissimilarity between two species, s_i and s_j, is conceptualized by a distance function, $d : S \times S \to \mathbb{R}_+$. In general, a distance function has the following properties. It is non-negative and symmetric, i.e. $d(s_i, s_j) = d(s_j, s_i) > 0$ for all $s_i, s_j \in S$ and $s_i \neq s_j$. Furthermore, $d(s_i, s_i) = 0$ for all $s_i \in S$, which expresses the very nature of what one means by 'dissimilarity': a species compared to itself does not have any dissimilarity.[14] The pairwise distances of all species are the elementary data upon which the diversity measure builds. Weitzman (1992, 1993) suggests the use of phylogenetic information to determine the pairwise distances between species, but also states that any other quantifiable trait of species could be used for that purpose as well, e.g. morphological or functional features. A distance function can, of course, also be meaningfully defined when species differ in more than one feature, for instance, as a weighted sum of differences in different features.

With given pairwise distances between all species, Weitzman's (1992) di-

---

[13]Solow et al. (1993) and Weikard (1998a, 1999, 2002) have developed biodiversity indices that follow a very similar logic.

[14]Sometimes the so-called triangle inequality, $d(s_i, s_j) \leq d(s_i, s_k) + d(s_k, s_j)$ for all $s_i, s_j, s_k \in S$, is invoked in addition to obtain a metric distance measure (e.g. Weikard 1998a, 1999, 2002). This is not necessary for developing the Weitzman index.

versity index $D(S)$ of a set S of species is then defined recursively by

$$\begin{aligned} D^W(Q \cup \{s_i\}) &= D^W(Q) + \delta(s_i, Q) \text{ for all } s_i \in S \backslash Q \text{ and } 0 \subset Q \subset S \quad (8.9)\\ &\text{where} \quad D^W(\{s_j\}) = D_0 \in \mathrm{I\!R}_+ \text{ for all } s_j \in S \\ &\text{and} \quad \delta(s_i, Q) = \min_{s_j \in Q} d(s_i, s_j) \text{ for all } s_i \in S \backslash Q \ . \end{aligned}$$

This means, the calculation of the index starts from an arbitrarily chosen start value $D_0 \in \mathrm{I\!R}_+$ assigned to the set that contains only one species, irrespective of what species s_j that is. Depending on the particular application, D_0 may be chosen to be zero or a very large number. One then calculates the biodiversity index of an enlarged subset Q' of S (with $0 \subset Q' \subseteq S$) that one obtains when adding species $s_i \in S \backslash Q$ to the set Q, $Q' = Q \cup \{s_i\}$, by adding the increase in diversity $\delta(s_i, Q)$ which species s_i adds to the diversity of the subset Q. This increase in diversity is calculated as the minimal distance between the added species s_i and any of the species s_j in the subset Q.[15] So, the recursive algorithm (8.9) allows one to calculate the diversity of a set S of species, starting from the arbitrarily chosen diversity value of a single species set, D_0, and then adding one species after the other until the whole set S is complete.

In general, the recursive algorithm (8.9) is path dependent, i.e. the value calculated for D^W depends on the particular sequence in which species are added when constructing the full set S. Therefore, the diversity function D^W as defined by Equation (8.9) is, in general, not unique.[16] However, Weitzman's diversity measure (8.9) is unique, and therefore most appealing, for the special case when the feature space is ultrametric.[17] Ultrametric spaces have an interesting geometric property which is also ecologically relevant: A set S of species characterized by ultrametric distances can be represented graphically by a hierarchical (e.g. phylogenetic) tree, and any hierarchical (phylogenetic) tree can be represented by ultramteric distances. Figure 8.2 shows an example of such a phylogenetic tree. In a phylogenetic tree, the distance $d(s_i, s_j)$, which indicates the dissimilarity between species s_i and s_j, is given by the vertical distance to the last common ancestor of s_i and s_j, and the diversity $D^W(S)$ of the set S of all species is given by the summed vertical length of all branches of the tree.

---

[15] This corresponds to the standard topological definition of the distance between a point and a set of points

[16] By imposing a condition called 'monotonicity in species', Weitzman (1992) can show that the class of, in general, path dependent diversity indices (8.9) reduces to a unique, path independent index which is given by $D(S) = \max_{s_i \in S}[D(S \backslash \{s_i\}) + \delta(s_i, S \backslash \{s_i\})]$.

[17] A space is called *ultrametric* if the pairwise distances between any three points in space have the property that the two greatest distances are equal: $\max\{d(s_i, s_j), d(s_j, s_k), d(s_i, s_k)\} = \mathrm{mid}\{d(s_i, s_j), d(s_j, s_k), d(s_i, s_k)\}$ for all $s_i, s_j, s_k \in S$.

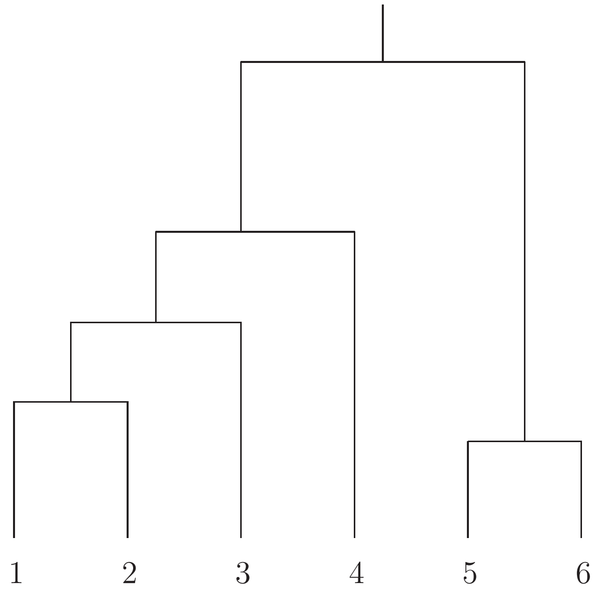

Figure 8.2 Phylogenetic tree representation of a set of six species with ultrametric distances (from Weitzman 1992: 370).

Nehring-Puppe index

Even more general than Weitzman's distance-function-approach is the so-called 'multi-attribute approach' proposed by Nehring and Puppe (2002, 2004). Like Weitzman, they base a measure of species diversity on the characteristic features of species. In contrast to Weitzman, the elementary data are not the pairwise dissimilarities between species, but the characteristic features f themselves. From the different features f and their relative weights $\lambda_f \geq 0$, which may be derived from the individuals' or society's preferences, Nehring and Puppe construct a diversity index as follows:

$$D^{NP}(\Omega) = \sum_{f \in F:\ \exists s_i \in S \text{ with '}s_i \text{ possesses feature } f\text{'}} \lambda_f. \tag{8.10}$$

In words, the diversity index for a set S of species is the sum of weights λ_f of all features f that are represented by at least one species s_i in the system. Each feature shows up in the sum at most once. In particular, each species s_i contributes to the diversity of the set S exactly the relative weight of all those features which are possessed by s_i and not already possessed by any other species in the set.

Nehring and Puppe also show that under certain conditions the characterization of an ecosystem by its diversity D^{NP} uniquely determines the relative weights λ_f of the different features. This means, in assigning a certain diversity to an ecosystem one automatically reveals an (implicit) value judgement about the relevant features according to which one distinguishes between species and one describes an ecosystem as more or less diverse.

8.4 Conceptual Comparison and Assessment

8.4.1 Diversity Indices and Value Judgments

All the diversity indices discussed here build – more or less obviously – on prior value judgements at different levels (Baumgärtner 2006). Technically, these value judgements enter the construction of the index as some parameter or underlying metric that determines how much weight is given to what information in calculating the index.

As for the ecological indices (cf. Section 8.3.2), the parameter α plays such a role: it determines how much weight is given to the unevenness of the distribution of relative abundances as compared to the weight of pure species number. The value of α is a priori arbitrary. In particular, it cannot be inferred from any ecological information about the system to be studied. So, there is no 'true' or 'correct' value of α, but its value has to be chosen by the scientist describing and analyzing a system for a particular purpose. For example, such a purpose may be to assess the development over time of a nature reserve in terms of its biodiversity and its associated potential to sustain ecosystem functioning; or it may be to compare two patches of rainforest in terms of their biodiversity of potentially useful pharmaceutical substances. Of course, with a particular purpose in mind, some values of α may be found to be better suited than others. But this choice of α then reflects individual or social preferences about why biodiversity is useful.

Similarly, in the Nehring-Puppe measure (cf. Section 8.3.3) it is the parameters λ_f which play this role: they determine the relative weight that the different features have in constructing the biodiversity index. For example, if pharmaceutical effectiveness is held to be a very important property of species, and their complementarity in the use of ecological resources is a less important property, a sample of species that differ mainly in pharmaceutical respect will be found to be more diverse than a sample of species differing mainly in ecological respect. And the result will be exactly the opposite if pharmaceutical effectiveness is taken to be a less important property than ecological complementarity. Again, the construction of the index, and the result it yields in terms of the level of biodiversity, depends on individual or social value judgments about why biodiversity matters.

This also applies to Weizman's index (cf. Section 8.3.3). Here, the weighting based on value judgments lies in the choice of a particular metric that is used for specifying the pairwise distances between species. For example, if species differ in two (or more) features, say pharmaceutical effectiveness and ecological complementarity, then it is the underlying metric that determines how the two dimensional description of species translates into one dimensional distances between them, which are then used to calculate the index. Weitzman claims that the pairwise distances between species are basic ecological information which comes, in principle, from ecological research. But since the metric introduces a weighting of different features, it involves individual or social value judgments and, therefore, is not purely ecological information. Like all other measures of biodiversity, the Weitzman index thus depends on prior judgments about why biodiversity is valuable.

8.4.2 Information Used and Not Used

Comparing the ecological and economic biodiversity indices (cf. Section 8.3) at the conceptual level, it is obvious that the two classes are distinct by the information they use for constructing a diversity index (Figure 8.3). While

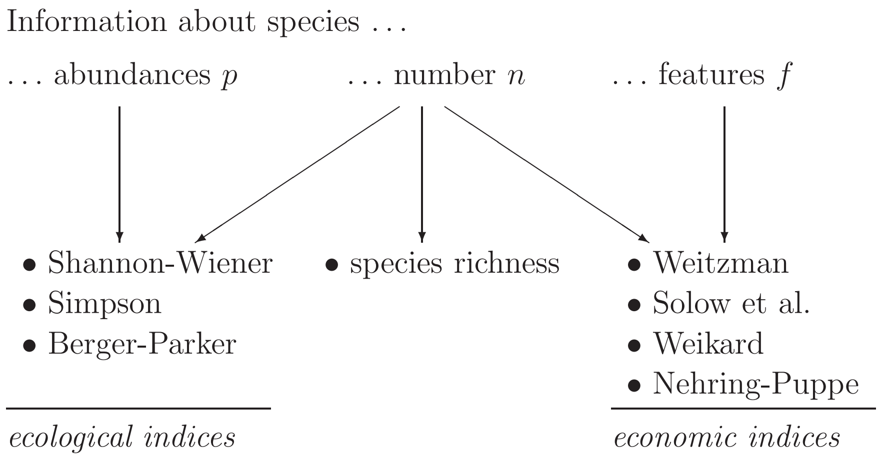

Figure 8.3 Biodiversity indices differ by the information on species and ecosystem composition they use.

the ecological measures (Section 8.3.2) use the number n of different species in a system as well as their relative abundances p, the economic ones (Section 8.3.3) use the number n of different species as well as their characteristic features f. In a sense, the indices discussed in Section 8.3.2 above measure 'heterogeneity' rather than 'diversity' (Good 1953, Hurlbert 1971, Peet 1974), as they are based on richness and abundances but completely miss out features. The indices discussed in in Section 8.3.3 above measure 'dissimilarity' rather than 'diversity', as they are based on richness and dissimilarity but completely miss out abundances. Both kinds of indices contain pure species richness as a special case.

Up to now, there do not exist any encompassing diversity indices based on all ecological information considered here – species richness n, abundances a, and features f. A logical next step at this point could be to construct a general diversity index based on species richness, abundances and features, which contains the existing indices as special cases. However, one should not jump to this conclusion too quickly. It is important to note that the ecological and economic diversity indices have come out of very different modes of thinking. They have been developed for different purposes and are based on fundamentally different value systems. Therefore, they may not even be compatible. This point is addressed in detail in the following.

8.4.3 Diversity of What? – The Relevance of Abundances and Features

From an economic point of view, relative abundances are usually considered irrelevant for the measurement of diversity. The reason is that in economics the diversity issue is usually framed as a choice problem. Diversity is then a property of the choice set, i.e. the set of feasible alternatives to choose from. Individuals facing a situation of choice should consider only the list of possible alternatives (say, the menu in a restaurant), rather than the actual allocation which has been realized as the result of other people's earlier choices (say, the dishes on the other tables in a restaurant). Furthermore, when economists talk about product diversity, relative abundances are irrelevant since there is the possibility of production.[18] If all people in a restaurant order the same dish from the menu, then this dish will be produced in the quantity demanded; and if all people order different dishes, then different dishes are produced. In any case, the diversity of the choice set is determined by the diversity of the order list (the menu), and not by the actual allocation of products (the dishes on the tables).

This argument has influenced economists view on biodiversity as well. Economists consider biological diversity as a form of product diversity, i.e. a diverse resource pool from which one can choose the most preferred option(s) (cf. Section 7.2). And this diversity is essentially determined by the choice set, i.e. the list S of species existent in an ecosystem (e.g. Weitzman 1992, 1993, 1998). The actual abundances of individuals of different species, in that view, do not matter.

Ecologists, in contrast, often argue that biological species living in natural ecosystems – even when considered merely as a resource pool to choose from – are different from normal economic goods for a number of reasons (e.g. Begon et al. 1998, Ricklefs and Miller 2000). First, individuals of a particular species

[18] While the scarcity of production factors may limit the *absolute* abundances of the produced products, all possible *relative* abundances can be produced without restriction.

cannot simply be produced; at least not so easily, not for any species, and not in any given number. Second, there are direct interactions between individuals and species within ecosystems, which heavily influence survival probabilities and dynamics in an ecosystem. And for that sake, relative abundances matter. And third, while some potential systems (in the sense of: relative abundance distributions) are viable in situ, others are not. For example, a community with very high relative abundance of predator species and very low relative abundance of prey species will go extinct altogether once the prey has been completely exhausted.

Hence, it becomes apparent that the two types of biodiversity measures – the ecological ones and the economic ones – aim at characterizing two very different aspects of the ecosystem. While the ecological measures describe the actual, and potentially unevenly distributed allocation Ω of species, the economic measures characterize the abstract list S of species existent in the system. In a sense, the two are not different measures of the same concept, but measures of two different concepts.

8.4.4 Diversity for What Purpose? – Different Philosophical Perspectives on Diversity

The underlying reason for this difference between the ecological and economic measures of biodiversity can be found in the philosophically distinct perspective on diversity between ecologists and economists. Ecologists traditionally view diversity more or less in what may be called a 'conservative' perspective, while economists predominantly have what may be called a 'liberal' perspective on diversity (Kirchhoff and Trepl 2001).

In the conservative view, which goes back to Gottfried Wilhelm Leibniz (1646–1716) and Immanuel Kant (1724–1804), diversity is an expression of unity. By viewing a system as diverse, one stresses the integrity and functioning of the entire system. The ultimate concern is with the system at large. In this view, diversity may have an indirect value in that it contributes to certain overall system properties, such as stability, productivity or resilience at the system level. In contrast, in the liberal view, which goes back to René Descartes (1596–1650), John Locke (1632–1704) and David Hume (1711–1776), diversity enables the freedom of choice for autonomous individuals who choose from a set of diverse alternatives. The ultimate concern is with the well-being of individuals. In this view, diversity of a choice set has a direct value in that it allows individuals to make a choice that better satisfies their individual subjective preferences. Once one alternative has been chosen, the other alternatives, and the diversity of the choice set, are no longer relevant.

Of course, the integrity and functioning of the entire system will also be important for the well being of autonomous individuals who simply want to choose from a set of diverse alternatives. For example, today's choice may impede the system's ability to properly work in the future and, therefore, to provide diversity to choose from in the future. This is an intertemporal argument, which combines (i) an argument about diversity's importance at a given point in time for individuals, who want to make an optimal choice at this point in time, and (ii) an argument about diversity's role for system functioning and evolution over time. From an analytical point of view, one should distinguish these two arguments. This underlies the distinction between the conservative and the liberal perspective, which is analytical to start with.

These two distinct perspectives on diversity – the conservative one and the liberal one – correspond to some extent with the two types of biodiversity measures considered here (Section 8.3): the ecological measures that take into account relative abundances, and the economic measures that deliberately do not take into account relative abundances. The ecological measures are based on a conservative perspective in that their main interest is to represent biodiversity as an indicator of ecosystem integrity and functioning. With that concern, the distribution of relative abundances is an essential ingredient in constructing a biodiversity index. In contrast, the economic measures are based on a liberal perspective in that their main interest is to represent biodiversity as a property of the choice set from which economic agents – individuals, firms or society – can choose to best satisfy their preferences. With that concern, it seems plausible that the actual distribution of relative abundances is not taken into account when constructing a biodiversity index.

8.5 Summary and Conclusion

I have reviewed the two broad classes of biodiversity measures currently being used, the ecological ones and the economic ones, and compared them at a conceptual level. It has turned out that the two classes are distinct by the information they use for constructing a diversity index. While the ecological measures use the number of different species in a system as well as their relative abundances, the economic ones use the number of different species as well as their characteristic features. Thereby, the two types of measures aim at characterizing two very different aspects of the ecosystem. The economic measures characterize the abstract list of species existent in the system, while the ecological measures describe the actual, and potentially unevenly distributed allocation of species.

I have argued that the underlying reason for this difference is in the philosophically distinct perspective on diversity between ecologists and economists. Ecologists traditionally view diversity more or less in what may be called a conservative perspective, while economists predominantly adopt what may be called a liberal perspective on diversity (Kirchhoff and Trepl 2001). In the former, the ultimate concern is with the integrity and functioning of a diverse system at large, while in the latter, the ultimate concern is with the well-being of individuals who want to make an optimal choice from a diverse resource base.

This difference in the philosophical perspective on diversity leads to using different information when constructing a measure of diversity. In the conservative perspective, the aim is to represent biodiversity as an indicator of ecosystem integrity and functioning. For tat purpose, the relative abundances of species are an important ingredient into a measure of biodiversity. In contrast, in the liberal perspective the aim is to represent biodiversity as a property of the choice set from which economic agents can choose to best satisfy their preferences. For that purpose, the characteristic features of species are very important, but relative abundances are not.

Hence, the question of how to measure biodiversity is intimately linked to the question of what is biodiversity good for (Baumgärtner 2006). This is not a purely descriptive question, but also a normative one. There are many possible answers, but in any case an answer requires value judgements. Do we consider biodiversity as valuable because it contributes to overall ecosystem functioning – either out of a concern for conserving the working basis of natural evolution, or out of a concern for conserving certain essential and life-supporting ecosystem services, such as oxygen production, climate stabilization, soil regeneration, and nutrient cycling (Barbier et al. 1994, Perrings et al. 1995a, Daily 1997b, Millennium Ecosystem Assessment 2005)? Or do we consider biodiversity as valuable because it allows individuals to make an optimal choice from a diverse resource base, e.g. when choosing certain desired genetic properties in plants for developing pharmaceutical substances (Polasky and Solow 1995, Polasky et al. 1993, Simpson et al. 1996, Rausser and Small 2000), or breeding or genetically engineering new food plants (Myers 1983, 1989, Plotkin 1988)?

These are examples for different value statements about biodiversity which are made on the basis of different fundamental value judgements: in the former case dominates the conservative perspective, in the latter the liberal one. As I have shown here, these two perspectives lead to different measures of biodiversity, the ecological measures and the economic measures. Of course, there is a continuous spectrum in between these two extreme views on why biodiversity is valuable and how to measure it. But in any case, one is lead to conclude, the measurement of biodiversity requires prior value judgments as to what purpose biodiversity serves in ecological-economic systems.

9. The Insurance Value of Biodiversity in the Provision of Ecosystem Services*

9.1 Introduction

In the face of uncertainty, diversity provides insurance for risk averse economic agents. For example, investors in financial markets diversify their asset portfolio in order to hedge their risk; firms diversify their activities, products or services when facing an uncertain market environment; farmers traditionally grow a variety of crops in order to decrease the adverse impact of uncertain environmental and market conditions. In this chapter, I argue that biological diversity plays a similar role: it can be interpreted as an insurance against the uncertain provision of ecosystem services, such as biomass production, control of water run-off, pollination, control of pests and diseases, nitrogen fixation, soil regeneration etc. Such ecosystem services are generated by ecosystems and are used by utility-maximizing and risk averse economic agents (Daily 1997b, Millennium Ecosystem Assessment 2005).[1]

In order to explore the hypothesis that biodiversity has an insurance value in the provision of ecosystem services, I take an interdisciplinary approach and study a conceptual ecological-economic model that combines (i) current results from ecology about the relationships between biodiversity, ecosystem functioning, and the provision of ecosystem services with (ii) economic methods to study decision-making of risk averse agents under uncertainty. The focus here is on how to model the ecology-economy-interface. Relevant economic and policy questions that arise from this view on biodiversity are only briefly sketched and are discussed in more detail elsewhere (Baumgärtner and Quaas 2005, Quaas and Baumgärtner 2006, Quaas et al. 2004).

Although ecologists usually stress the large extent of ignorance about the detailed mechanisms of ecosystem functioning (e.g. Holling et al. 1995, Loreau et al. 2001, Schulze and Mooney 1993), there now seems to be a consensus about some of the basic mechanisms through which biodiversity influences ecosystem

*Forthcoming in *Natural Resource Modeling*.

[1]See the detailed discussion in Chapter 7 (p. 129) on biodiversity and ecosystem services.

functioning and the provision of ecosystem services (Hooper et al. 2005, Kinzig et al. 2002, Loreau et al. 2001, 2002b). Among other insights, it has become clear that biodiversity may decrease the variability of these services. This result has lead economists to suggest that biodiversity may be seen as a form of insurance, for instance in agriculture or medicine (Perrings 1995a, Schläpfer et al. 2002, Swanson and Goeschl 2003, Weitzman 2000). On the other hand, availability of financial insurance against the over- or under-provision with ecosystem services, or other financial products that allow the hedging of income risk, may be seen as substitutes for the natural insurance provided by biodiversity (Quaas and Baumgärtner 2006, Ehrlich and Becker 1972). The implications of this idea for both economic well-being and the state of ecosystems in terms of biodiversity, however, have hardly been explored so far.

One notable exception is to be found in the field of agricultural economics. A number of studies have analyzed the contribution of crop diversity to the mean and variance of agricultural yields (Smale et al. 1998, Schläpfer et al. 2002, Widawsky and Rozelle 1998, Zhu et al. 2000) and to the mean and variance of farm income (Di Falco and Perrings 2003, 2005, Di Falco et al. 2005). It has been conjectured that risk averse farmers use crop diversity in order to hedge their income risk (Birol et al. 2005a, 2005b, Di Falco and Perrings 2003) and that this may be affected by agricultural policies such as subsidized crop yield insurance or direct financial assistance (Di Falco and Perrings, 2005).[2]

With this analysis, I want to look into these issues in greater generality and with a particular focus on modeling the ecology-economy interface. In order to study the role of biodiversity as a form of natural insurance I employ a conceptual model that captures the relevant ecological and economic relationships in a stylized way. While such a simple model cannot offer any quantitative predictions or detailed policy prescriptions, it can clarify the underlying theoretical structure of the problem: The ecosystem generates a valuable ecosystem service at a level that is uncertain because of environmental stochasticity. Its probability distribution is influenced by the level of biodiversity, which is measured by a suitable index. In line with evidence from ecology, I posit a monotonically increasing and concave relationship between biodiversity and the mean absolute level of the ecosystem service provided by the ecosystem, and a monotonically decreasing and convex relationship between biodiversity and the variance of ecosystem service. The ecosystem service is being used by an ecosystem manager, say, a farmer, who is assumed to be a risk averse expected utility maximizer. Protection of biodiversity is costly. There exists a financial form of insurance against over- or under-provision with the ecosystem service. The ecosystem manager decides upon (i) the level of biodiversity and (ii) the level

[2]In this respect, biodiversity plays a similar role for farmers as other risk changing production factors, such as e.g. nitrogen fertilizer or pesticides (Horowitz and Lichtenberg 1993, 1994a, b).

of financial insurance coverage.

In this framework, I analyze the optimal allocation of biodiversity as a choice of endogenous environmental risk in mean-variance space.[3] In particular, I

- determine the insurance value of biodiversity, i.e. the marginal value of biodiversity in its function to reduce the risk premium of the ecosystem manager's income risk from using ecosystem services under uncertainty,
- study the optimal allocation of funds in the trade-off between investing into natural capital, that is, biodiversity protection, and the purchase of financial insurance, and
- analyze the effect of different institutional regimes in the market for financial insurance (e.g. availability, transaction costs and profitability of financial insurance) on biodiversity protection.

I conclude that biodiversity acts as a form of natural insurance for risk averse ecosystem managers against the over- or under-provision with ecosystem services. Therefore, biodiversity has an insurance value, which is a value component in addition to the usual value arguments (such as direct or indirect use or non-use values, or existence values)[4] which hold in a world of certainty. In this respect, biodiversity and financial insurance are substitutes. Hence, the availability, and the exact institutional design, of financial insurance, influence the level of biodiversity protection.

The chapter is organized as follows. Section 9.2 discusses the ecological background and surveys the relevant literature. Section 9.3 introduces a formal ecological-economic model. The model analysis and results are presented in Section 9.4, with all formal derivations and proofs given in the Appendix. Section 9.5 critically discusses the limitations and the generality of the results, and Section 9.6 concludes.

9.2 Ecological Background: Biodiversity and the Provision of Ecosystem Services

Over the past fifteen years, there has been intensive research in ecology on the role of biodiversity for ecosystem functioning and the provision of ecosystem

[3]This procedure has been inspired by Crocker and Shogren (1999, 2001, 2003) and Shogren and Crocker (1999). It is also employed by Baumgärtner and Quaas (2005) and Quaas and Baumgärtner (2006).

[4]See the detailed discussion on the economic value of biodiversity in Section 7.3.

services. 'Biodiversity' has been defined as 'the variability among living organisms from all sources [...] and the ecological complexes of which they are part' (CBD 1992), which encompasses a wide spectrum of biotic scales, from genetic variation within species to biome distribution on the planet (Gaston 1996, Purvis and Hector 2000, Wilson 1992). Biodiversity can be described in terms of numbers of entities (e.g. genotypes, species, or ecosystems), the evenness of their distribution, the differences in their functional traits, and their interactions. The simplest measure of biodiversity at, say, the species level is therefore simply the number of different species ('species richness'). Much of ecological research has relied on this measure when quantifying 'biodiversity', although more encompassing information has also been employed.[5]

Research on the role of biodiversity for ecosystem functioning and the provision of ecosystem services built on (i) observations of existing ecosystems, (ii) controlled experiments both in the laboratory and in the field ('pots and plots') and (iii) theory and model analysis. While the discussion of results has been, at times, heated and controversial, there now seems to be a consensus over some of the basic results from this research (Hooper et al. 2005, Kinzig et al. 2002, Loreau et al. 2001, 2002b).[6] Among other insights two 'stylized facts' about biodiversity and ecosystem functioning emerged which are of crucial importance for the issue studied here:

1. *Biodiversity may enhance ecosystem productivity.* In many instances, an increase in the level of biodiversity monotonically increases the mean absolute level at which certain ecosystem services (e.g. biomass production or nutrient retention) are provided. This effect decreases in magnitude with the level of biodiversity.

2. *Biodiversity may enhance ecosystem stability.* In many instances, an increase in the level of biodiversity monotonically decreases the temporal variability of the level at which these ecosystem services are provided under changing environmental conditions. This effect decreases in magnitude with the level of biodiversity.

These two stylized facts are now discussed in turn.[7]

---

[5]The question of how to construct an aggregate and encompassing measure of biodiversity has been extensively discussed and is still subject to on-going research (Baumgärtner 2004b, Crozier 1992, Magurran 2004, May 1990, Nehring and Puppe 2004, Peet 1974, Purvis and Hector 2000, Vane-Wright 1991, Weitzman 1992, 1998, Whittaker 1972). See the detailed discussion of this issue in Chapter 8.

[6]The article by Hooper et al. (2005) is a committee report commissioned by the Governing Board of the Ecological Society of America. Some of its authors have previously been on opposite sides of the debate. This report surveys the relevant literature, identifies a consensus of current knowledge as well as open questions, and can be taken to represent the best currently available ecological knowledge about biodiversity and ecosystem functioning.

[7]This discussion is compiled from the report of Hooper et al. (2005: Sections II.A and

9.2.1 Biodiversity May Enhance Ecosystem Productivity

The absolute level of a certain ecosystem service (e.g. biomass production, carbon sequestration or nitrogen fixation) may be influenced by species or functional diversity in several ways.[8] Indeed, more than 50 potential response patterns have been proposed (Loreau 1998a, Naeem 2002). There are two primary mechanisms through which biodiversity may increase the mean absolute level at which certain ecosystem services are provided (Figure 9.1):

(i) Only one or a few species might have a large effect on any given ecosystem service. Increasing species richness, i.e. the number of different species, increases the likelihood that those key species would be present in the system (Aarssen 1997, Huston 1997, Loreau 2000, Tilman et al. 1997b). This is known as the 'sampling effect' or the 'selection probability effect' (Figure 9.1A).[9]

(ii) Species or functional richness could increase the level of ecosystem services through complementarity – i.e. species use different resources, or the same resources but at different times or different points in space – and facilitation – i.e. positive interactions among species so that e.g. certain species alleviate harsh environmental conditions or provide a critical resource for other species. Both complementarity and facilitation lead to an 'overyielding effect' (Figure 9.1B), in which biomass production in mixtures exceeds expectations based on monoculture yields (Ewel 1986, Harper 1977, Hector et al. 1999, Loreau 1998b, Trenbath 1974, Vandermeer 1989).

Complementarity, facilitation and sampling effects will all lead to a saturating average impact of species richness on the level of some ecosystem service (Figure 9.1A, B).

Experiments have confirmed the important role of these two primary mechanisms through which biodiversity may increase the mean absolute level of certain ecosystem services. This holds, in particular, for experiments with herbaceous plants, in which average primary production and nutrient retention were found to increase with increasing plant species or functional richness, at

---

II.B), with large parts being original quotes from this report. For a more detailed and encompassing discussion see Hooper et al. (2005).

[8] The patterns depend on the degree of dominance of the species lost or gained, the strength of their interactions with other species, the order in which species are lost, the functional traits of both the species lost and those remaining, and the relative amount of biotic and abiotic control over process rates (Lawton 1994, Naeem 1998, Naeem et al. 1995, Sala et al. 1996, Vitousek and Hooper 1993).

[9] There is still disagreement over whether sampling effects are relevant to natural ecosystems, or whether they only occur in artificially assembled systems (Huston 1997, Loreau 2000, Mouquet et al. 2002, Schläpfer et al. 2005, Tilman et al. 1997b, Wardle 1999).

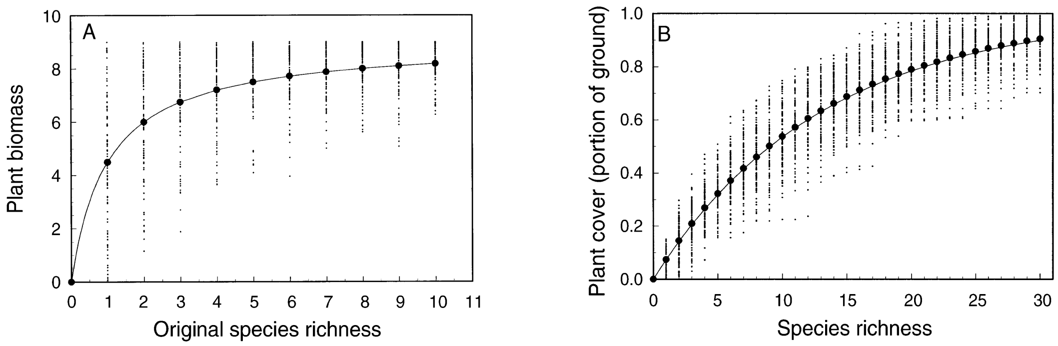

Figure 9.1 Ecological theory has suggested two basic mechanisms of how biodiversity could increase the mean absolute level of ecosystem services: sampling or selection probability effect (A), and complementarity or facilitation (B). Points show individual treatments, and lines show the average response. (Figures are taken from Tilman 1997b, as compiled by Hooper et al. 2005.)

least within the range of species richness tested and over the relatively short duration of the experiments (Fridley 2003, Hector et al. 1999, Loreau and Hector 2001, Niklaus et al. 2001, Reich et al. 2001, Tilman et al. 1996, 1997a, 2001, 2002).[10] In these experiments, the responses to changing diversity are strongest at low levels of species richness and generally saturate at 5–10 species. It has also become evident that complementarity, facilitation and sampling/selection effects are all relevant and can be observed in experiments.[11] They are not necessarily mutually exclusive, but they may be simultaneously or sequentially at work in one system. The strength of species complementarity and interspecific facilitation and, thus, the quantitative response in the level of ecosystem services to changes in species richness varies with both the functional characteristics of the species involved and the biotic as well as abiotic environmental context.

These general findings need to be qualified in a number of respects:

- Experiments have shown that the exact response of ecosystem services on changes in biodiversity is determined at least as much by differences in

---

[10]Much of the experimental work has focused on the effect of plant diversity on primary production and nutrient retention. Recently, evidence for ecosystem services other than biomass production and from ecosystems other than grasslands has begun to accumulate as well. Important insights come from research on intercropping and agroforestry (Ewel 1986, Fridley 2001, Harper 1977, Hector et al. 2002, Loreau 1998b, Smale et al. 1998, Trenbath 1974, Vandermeer 1990, Zhu et al. 2000).

[11]Identifying the exact mechanisms by which experimental manipulation of species leads to increased levels of ecosystem processes has led to substantial debate (Aarssen 1997, Garnier et al. 1997, Hector et al. 2000, Huston 1997, Huston and McBride 2002, Huston et al. 2000, Schmid et al. 2002, van der Heijden et al. 1999, Wardle 1999), as many experiments were originally designed to test general patterns, rather than to test the underlying mechanisms.

species composition, i.e. which species and functional traits are lost and remain behind, as by species richness, i.e. how many species are lost.

- Patterns of response to experimental manipulation of species richness vary for different ecosystem processes and services, different ecosystems, and even different compartments within ecosystems.

- Varying multitrophic diversity and composition, i.e. the diversity and composition of an ecological community at more than one trophic level, can lead to more idiosyncratic behavior than varying diversity of primary producers alone.

The different patterns found under experimental conditions may or may not reflect actual patterns seen for a particular ecosystem under a particular scenario of species loss or invasion, which will depend not only on the functional traits of the species involved, but also on the exact pattern of environmental change and the species traits that determine how species respond to changes in environmental conditions (Lavorel and Garnier 2002, Schläpfer et al. 2005, Symstad and Tilman 2001).

9.2.2 Biodiversity May Enhance Ecosystem Stability

The debate about whether (or not) biodiversity enhances ecosystem stability, i.e. whether (or not) ecosystem properties are more stable in response to environmental fluctuations as diversity increases, has a long tradition in ecology (McCann 2000). This so-called 'diversity-stability-debate' has been initiated in the 1950s by observations from natural ecosystems which were found to be more productive and more stable when more diverse (Elton 1958, Odum 1953, MacArthur 1955). This early diversity-stability-hypothesis has been shaken in the early 1970s by computer simulations of ecosystems which demonstrated that these systems were more *un*stable when more diverse (May 1972, 1974). However, because the simulated model systems were randomly and purely fictional, the diversity-stability-question for real ecosystems remained open.[12] In the 1990s, the debate gained new momentum and research was organized and discussed more systematically, with results coming from controlled laboratory experiments, field studies and theoretical analysis.

The diversity-stability-debate is generally clouded by inconsistent terminology, as 'stability' is an umbrella term that denotes a large number of potential

[12]The simulated model systems in the analysis of May (1972, 1974) were randomly constructed by putting together a given number of system elements (species) and, in particular, linking them by randomly assigned interaction strengths which were taken from a uniform distribution over all possible interaction strengths. This is in contrast to recent empirical evidence that in real ecosystems the vast majority of pairwise interactions are weak (Paine 1992, Wootton 1997, McCann et al. 1998).

phenomena, including, but not limited to, resistance to disturbance, resilience to disturbance, temporal variability in response to fluctuating abiotic conditions, and spatial variability in response to differences in either abiotic conditions or the biotic community (Chesson 2000, Chesson et al. 2002, Cottingham et al. 2001, Grimm and Wissel 1997, Holling 1986, Lehmann and Tilman 2000, Loreau et al. 2002a, May 1974, McNaughton 1993, Peterson et al. 1998, Pimm 1984). Most research so far has focused on temporal variability, but some of the results may also apply to other measures of ecosystem stability.

Theory, both via simple ecological reasoning and via mathematical models, has lead to the understanding that a diversity of species with different sensitivities to a suite of environmental conditions should lead to greater stability of ecosystem properties. The basic idea is that with increasing number of functionally different species, the probability increases that some of these species can react in a functionally differentiated manner to external disturbance of the system and changing environmental conditions. In addition, the probability increases that some species are functionally redundant, such that one species can take over the role of another species when the latter goes extinct. This is what ecologists have been calling an 'insurance effect' of biodiversity in carrying out ecological processes (Borrvall et al. 2000, Elton 1958, Chapin and Shaver 1985, Hooper et al. 2002, Lawton and Brown 1993, MacArthur 1955, Naeem 1998, Naeem and Li 1997, Petchey et al. 1999, Trenbath 1999, Walker 1992, Walker et al. 1999, Yachi and Loreau 1999).[13] With this logic, processes that are carried out by a relatively small number of species are hypothesized to be most sensitive to changes in diversity (Hooper et al. 1995). Also, the number of species or functional traits necessary to maintain ecosystem processes under changing environmental conditions increases with spatial and temporal scales (Casperson and Pacala 2001, Chesson et al. 2002, Field 1995, Pacala and Deutschman 1995).

Several mathematical models generally support these hypotheses (see McCann 2000, Cottingham et al. 2001, Loreau et al. 2002a for reviews) and highlight the role of statistical averaging – the so-called 'portfolio effect' – for the result (Doak et al. 1998, Tilman et al. 1998): if species abundances are negatively correlated or vary randomly and independently from one another, then overall ecosystem properties are likely to vary less in more diverse communities than in species-poor communities.[14] The strength of the modeled effects of di-

[13] In such cases, there is compensation among species: as some species do worse, others do better due to differences in their functional traits. As a result, unstable individual populations stabilize properties of the ecosystem as a whole. Hence, instability of the community composition is no contradiction to, but may actually support stability of ecosystem processes (Ernest and Brown 2001, Hughes and Roughgarden 1998, Ives et al. 1999, Landsberg 1999, Lehman and Tilman 2000, McNaughton 1977, Tilman 1996, 1999, Walker et al. 1999).

[14] This is similar to the effect of diversifying a portfolio of financial assets, e.g. stocks.

versity depends on many parameters, including the degree of correlation among different species' responses (Chesson et al. 2002, Doak et al. 1998, Lehman and Tilman 2000, Tilman 1999, Tilman et al. 1998, Yachi and Loreau 1999), the evenness of distribution among species' abundances (Doak et al. 1998), and the extent to which the variability in abundances scales with the mean (Cottingham et al. 2001, Tilman 1999, Yachi and Loreau 1999).[15]

While theory is well developed and predicts that increased diversity will lead to lower variability of ecosystem properties under those conditions in which species respond in a differentiated manner to variations in environmental conditions, it cannot tell us how important the underlying basic mechanisms are in the real world or whether they saturate at high or low levels of species richness. This requires experimental investigations. However, controlled experiments are very difficult to carry out, because one needs to make sure that the effect of species diversity is not confounded by other variables, such as e.g. soil fertility or disturbance regime. Nevertheless, considerable evidence exists from field studies in a variety of ecosystems that in diverse communities, redundancy of functional traits and compensation among species can buffer ecosystem processes in response to changing conditions and species loss. Examples include studies of arctic tundra (Chapin and Shaver 1985), Minnesota grasslands (Tilman 1996, 1999, Tilman et al. 2002), deserts (Ernest and Brown 2001), lakes (Frost et al. 1995, Schindler et al. 1986), and soil ecosystems (de Ruiter et al. 2002, Griffiths et al. 2000, Ingham et al. 1985, Liiri et al. 2002). As an example, Figure 9.2 shows experimental results for aboveground plant biomass production in response to climatic variability in a Minnesota grassland (Figure 9.2A), and net ecosystem CO_2 flux in a microbial microcosm (Figure 9.2B). While the overall stability patterns found are as predicted from theory, the experiments so far give little insights about the underlying basic mechanisms. Also, mechanisms other than compensation among species can affect stability in response to changing species richness.

Several experiments that manipulate diversity in the field and in microcosms generally support theoretical predictions that increasing species richness increases stability of ecosystem properties. For instance, stability of plant production, as measured by resistance and/or resilience to nutrient additions, drought and grazing, increased with the Shannon-Wiener index of diversity[16] in a variety of successional and herbivore-dominated grasslands (McNaughton 1977, 1985, 1993). And in Minnesota grasslands, resistance to loss of plant productivity to drought increased with increasing plant species richness (Tilman

---

[15] It is generally acknowledged that the underlying assumptions of the mathematical models as to these parameters need further investigation and more experimental confirmation. Also, the role of the stability measures used and other mechanisms built into the models (such as e.g. overyielding) need further clarification.

[16] See Footnote 5.

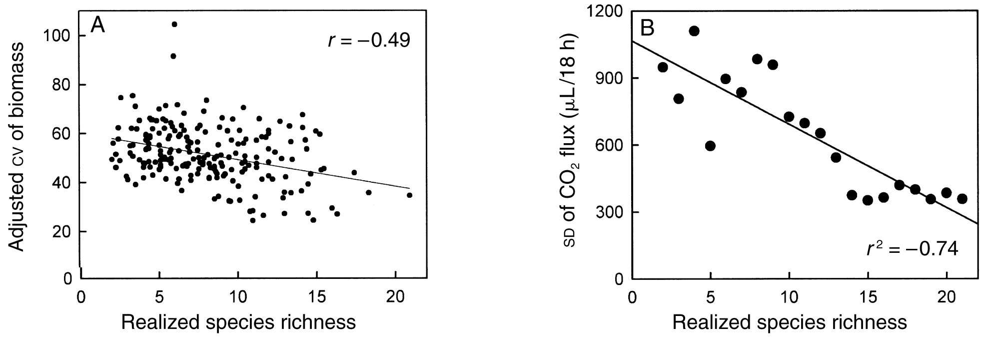

Figure 9.2 Ecological experiments found that species richness may decrease the variability of ecosystem services, such as e.g. aboveground plant biomass production in response to climatic variability in a Minnesota grassland (A), or net ecosystem CO_2 flux in a microbial microcosm (B). (Figures are taken from Tilman 1999 [A] and McGrady-Steed et al. 1997 [B], as compiled by Hooper et al. 2005 and Loreau et al. 2001.)

and Downing 1994). However, results of these experiments may be confounded by a variety of variables other than species richness or diversity, which has raised considerable controversy over the interpretation of these results (e.g. Givnish 1994, Grime 1997, Grime et al. 2000, Huston 1997, Huston et al. 2000, Pfisterer and Schmid 2002). Experiments in microcosms and grasslands suggest that increased species richness, either in terms of numbers of different functional groups, or numbers of species within trophic functional groups, can lead to decreased temporal variability in ecosystem properties (Emmerson et al. 2001, McGrady-Steed et al. 1997, Naeem and Li 1997, Petchey et al. 1999, Pfisterer et al. 2004; but see also Pfisterer and Schmid 2002). But while species richness or the Shannon-Wiener-index of species diversity was statistically significant in all these experiments, species composition (where investigated) had an at least equally strong effect on stability.

In sum, the experimental work provides qualified support for the hypothesis that species richness can increase the stability of ecosystem processes and services, although the underlying mechanisms can differ from theoretical predictions and in many cases still need to be fully resolved (Loreau et al. 2001).

9.3 Ecological-Economic Model

In order to study the economic implications of the insights from ecology about how biodiversity affects ecosystem functioning and the provision of ecosystem services, I shall cast them into a simple and stylized ecological-economic model.

9.3.1 Biodiversity and the Provision of Ecosystem Services

For notational simplicity, consider only one ecosystem service and let s be the amount generated of that service. As an example, think of insects providing pollination service to an orchard farmer. Because of environmental stochasticity the level s, at which the ecosystem service is provided, is a random variable. Assume, for analytical simplicity (and lack of specific ecological evidence on this point), that s is normally distributed with mean μ_s and standard deviation σ_s.

As discussed in the previous section, ecological research provides evidence that the level of biodiversity affects the statistical distribution of the ecosystem service. Let $v \in [0, \infty]$ be an appropriate index of biodiversity.[17] The two stylized facts about the relationship between biodiversity and the provision of ecosystem services, which emerged from ecological research (cf. Section 9.2), can then formally be expressed as:

$$\mu_s = \mu_s(v) \quad \text{with} \quad \mu_s'(v) > 0\ , \quad \mu_s''(v) \leq 0\ , \tag{9.1}$$

$$\sigma_s = \sigma_s(v) \quad \text{with} \quad \sigma_s'(v) < 0\ , \quad \sigma_s''(v) \geq 0\ , \tag{9.2}$$

where the prime denotes a derivative. That is, the mean level of ecosystem service increases and the standard deviation decreases with the level of biodiversity. Both effects decrease in magnitude with the level of biodiversity. While biodiversity, thus, is beneficial in a twofold manner – by increasing the mean level, at which the ecosystem service is being provided, and by decreasing its standard deviation – its provision is costly. Assume that the (direct and opportunity) costs of biodiversity are given by a cost function

$$C(v) \quad \text{with} \quad C'(v) > 0, \quad C''(v) \geq 0\ . \tag{9.3}$$

In the example of an orchard farmer using insects' pollination services, the costs of biodiversity could result from setting aside land from agricultural cultivation and leaving it in a natural state, so that hedges and wetlands can provide habitat for insects.[18]

---

[17] According to the discussion in the previous section, 'biodiversity' could in many instances simply be measured by the number of different species ('species richness'). However, the discussion in the previous section also suggests that in some instances it should be measured by a more sophisticated index which takes into account the functional traits and relative abundances of different species as well as their interactions (see Footnote 5).

[18] According to the well established species-area-relationships, the level of biodiversity v increases with the area l of land as $v \sim l^z$, where z (with $0 < z < 1$) is a characteristic constant for the type of ecosystem (MacArthur and Wilson 1967, Rosenzweig 1995, Gaston 2000). Assuming constant per-hectare-costs of land, this leads to a strictly convex cost function.

9.3.2 Ecosystem Manager

The ecosystem manager, who manages the system for the services s it provides, chooses the level of biodiversity $v \in [0, \infty]$.[19] On the one hand, the choice of v implies costs as given by Equation (9.3). On the other hand, biodiversity is essential for ecosystem functioning and the provision of ecosystem services. The ecosystem manager has benefits from ecosystem services, $B(s)$. For simplicity, assume that:

$$B(s) = s \ . \tag{9.4}$$

Since ecosystem services s are a random variable (normally distributed with mean μ_s and standard deviation σ_s) and the level of biodiversity v determines the distribution of this random variable according to (9.1) and (9.2), the benefits are also a random variable normally distributed with mean $\mu_s(v)$ and standard deviation $\sigma_s(v)$. The ecosystem manager's net income y is then given by

$$y = B(s) - C(v) = s - C(v) \ , \tag{9.5}$$

which is a random variable normally distributed with mean μ_y and standard deviation σ_y:

$$\mu_y(v) = \mu_s(v) - C(v) \quad \text{and} \tag{9.6}$$

$$\sigma_y(v) = \sigma_s(v) \ . \tag{9.7}$$

Hence, by choosing the level of biodiversity v, the ecosystem manager chooses a particular (normal) distribution $N(\mu_y(v), \sigma_y(v))$ of net income. That is, he chooses a particular income 'lottery' (Crocker and Shogren 2001).

The ecosystem manager's preferences over his uncertain net income y are represented by a von Neumann-Morgenstern expected utility function

$$U = \mathcal{E}[u(y)], \tag{9.8}$$

where $\mathcal{E}$ is the expectancy operator and $u(y)$ is a Bernoulli utility function which is assumed to be increasing ($u' > 0$) and strictly concave ($u'' < 0$), i.e. the ecosystem manager is non-satiated and risk averse.[20] In order to obtain

---

[19] Of course, it is a major simplification to assume that one can directly choose a certain level of biodiversity. Actually, one would choose some instrumental variable, such as area of protected land, or investment in some species protection/recovery plan, which then results in a certain level of biodiversity. Chapter 11 deals in more detail with the question of how to attain a certain level of biodiversity.

[20] While risk-aversion is a natural and standard assumption for farm *households* (Besley 1995, Dasgupta 1993: Chapter 8), it appears as an induced property in the behavior of (farm) *companies* which are fundamentally risk neutral but act as if they were risk averse when facing e.g. external financing constraints or bankruptcy costs (Caillaud et al. 2000, Mayers and Smith 1990).

simple closed-form solutions, assume that $u(y)$ is the constant absolute risk aversion Bernoulli utility function

$$u(y) = -e^{-\rho\, y}, \tag{9.9}$$

where $\rho > 0$ is a parameter describing the ecosystem manager's Arrow-Pratt measure of risk aversion (Arrow 1965, Pratt 1964). The ecosystem manager's von Neumann-Morgenstern expected utility function (9.8) is then given by (see Appendix A9.1)

$$U = \mu_y - \frac{\rho}{2}\, \sigma_y^2\ , \tag{9.10}$$

which is the simplest expected utility function of the mean-variance type.

9.3.3 Financial Insurance

In order to analyze the influence of availability of financial insurance products on the ecosystem manager's choice of biodiversity (in Section 9.4.4), financial insurance is introduced in a simple and stylized way.[21] I assume that the manager does or does not have the option of buying financial insurance under the following contract:

- The insurant chooses the fraction $a \in [0, 1]$ of insurance coverage.
- He receives (pays)

  $$a\,(s - \mu_s) \tag{9.11}$$

  from (to) the insurance company as an actuarially fair indemnification benefit (risk premium) if his realized income is below (above) the mean income.[22]
- In addition, he pays a mark-up for the transaction costs of insurance and the insurance company's profit:

  $$\frac{\delta}{2}\, a^2\, \sigma_s^2\ , \tag{9.12}$$

  where $\delta \geq 0$ is a parameter describing how actuarially unfair is the insurance contract. Thus, the costs of insurance over and above the actuarially

[21] This stylized insurance institution is a special case of the one studied by Quaas and Baumgärtner (2006).

[22] This benefit/premium-scheme is actuarially fair, because the insurance company has an expected net payment stream of $\mathcal{E}[a(s - \bar{s})] = 0$. To the insurant, this actuarially fair benefit/premium-scheme does not come at any real costs, as $\mathcal{E}[a(s - \bar{s})] = 0$. It is fully equivalent to the traditional model of insurance (e.g. Ehrlich and Becker 1972: 627) where losses compared with the maximum income are insured against and one pays a constant insurance premium irrespective of actual income.

fair risk premium, which are a measure of the 'real' costs of insurance to the insurant,[23] are assumed to follow a quadratic cost function.

This is a highly idealized form of financial insurance which captures in the most simple way the essence of financial insurance with an actuarially fair risk premium and some mark-up (due to transaction costs and profits) on top. The higher the insurance coverage a, the lower the effective income risk; and the effective income risk can be continuously reduced down to zero by increasing a to one. This follows the 'Venetian Merchant'-model of insurance: there exists an insurance company (the 'Venetian Merchant') which is ready to (fully or partially) take over the income risk from the insurant. In order to abstract from any problems related to informational asymmetry I assume that the statistical distribution $N(\mu_s, \sigma_s)$ and actual level s of ecosystem service are observable to both insurant and insurance company.

9.4 Analysis and Results

When analyzing the insurance value of biodiversity (Section 9.4.2), the optimal allocation of biodiversity (Section 9.4.3), and the effect of different institutional settings in the market for financial insurance products on biodiversity protection (Section 9.4.4), the idea is to treat the level of biodiversity v as the choice variable and to analyze the choice of biodiversity as the choice of an income lottery.

9.4.1 The Choice Set

To start with, neglect the option to buy financial insurance and focus on biodiversity as the natural insurance. Financial insurance will be taken into account in Section 9.4.4. As v can range from zero to infinity, the resulting feasible and efficient distributions of net income y (Equation 9.5) in μ_y-σ_y-space can be depicted by an income-possibility-frontier as in Figure 9.3. Income distributions above the income-possibility-frontier are not feasible; income distributions below the income-possibility-frontier may be feasible, but are not efficient.

The right hand end of the curve corresponds to very low levels of biodiversity v: the standard deviation σ_y of income is high. As v increases, one moves left along the curve: the standard deviation of income is reduced due to the stabilizing effect of biodiversity (Equations 9.2 and 9.7) and the mean income increases, because the mean level of ecosystem service increases with

[23]Since the actuarially fair risk premium does not cause any expected payoff/costs to the insurant, only the price component over and above the actuarially fair risk premium (the so-called 'loading' of the premium) constitutes real costs of insurance to the insurant (Ehrlich and Becker 1972: 626-627).

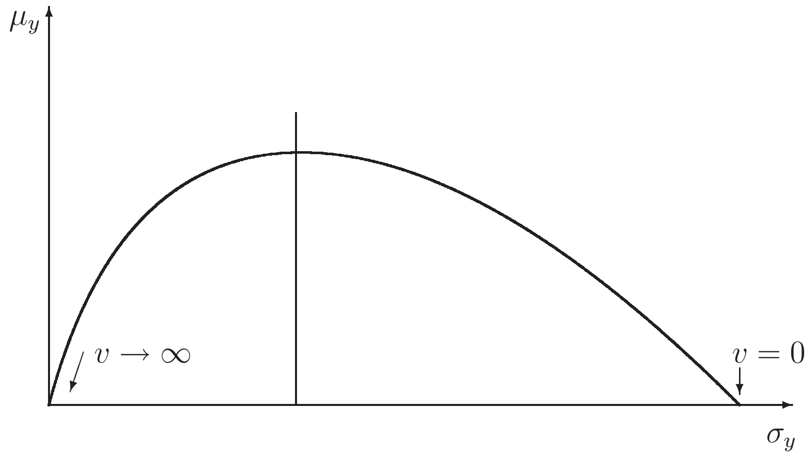

Figure 9.3 Feasible and efficient distributions of net income y (Equation 9.5) in μ_y-σ_y-space are represented by the income possibility frontier (solid line). The vertical line separates the domain with a trade-off between mean and standard deviation of income (left) from the domain without such a trade off (right).

biodiversity while the costs of biodiversity are not too important at low levels of biodiversity (Equations 9.1, 9.3 and 9.6). As the level v of biodiversity increases further, i.e. moving left along the curve even further, the additional mean benefits from additional ecosystem service become smaller and smaller (Equation 9.1) while the additional costs of biodiversity become greater and greater (Equation 9.3), thus eventually causing additional mean net benefits y from biodiversity to become negative. This corresponds to the left hand end of the curve: as biodiversity v increases (i.e. moving left along the curve) the standard deviation σ_y of income still decreases while the mean income μ_y decreases.

Overall, the income possibility frontier in μ_y-σ_y-space has two parts: in the left hand part (corresponding to high levels v of biodiversity) the mean income μ_y increases with increasing standard deviation σ_y; in the right hand part (corresponding to low levels v of biodiversity) the mean income μ_y decreases with increasing standard deviation σ_y. Given the ecosystem manager's expected utility function (9.10), according to which a high mean income and a low standard deviation of income are desirable, this means that for low levels of biodiversity there does not exist any economic problem. For, increasing the level of biodiversity at low v (right hand part of the curve) has a double desirable effect: it increases the mean income and it reduces the standard deviation of income. In contrast, for high levels of biodiversity (left hand part of the curve) when (opportunity) costs of biodiversity become important, the ecosystem manager faces a trade-off: increasing the level of biodiversity reduces the standard deviation of income, but reduces mean income, too.

It is the left part of the curve which suggests the interpretation that biodiversity provides insurance. As with buying financial insurance, increasing the level of biodiversity reduces the standard deviation of income, but reduces mean income, too. In this domain, a choice has to be made in order to optimally balance the two opposing goals of a high mean income and a low standard deviation of income.

9.4.2 The Insurance Value of Biodiversity

In order to precisely define the insurance value of biodiversity, let me come back to the idea that the ecosystem can be seen as an infinite set of lotteries (Crocker and Shogren 2001). By choosing the level of biodiversity v, the ecosystem manager determines the distribution $N(\mu_s(v), \sigma_s(v))$ of ecosystem service (Equations 9.1 and 9.2), which then determines the distribution $N(\mu_y(v), \sigma_y(v))$ of income (Equations 9.6 and 9.7). Thus, by choosing the level of biodiversity v, he chooses a particular income lottery. In the model employed here, this lottery is uniquely characterized by the level of biodiversity v. Therefore, one may speak of 'the lottery v'.

One standard method of how to value the riskiness of a lottery to a decision maker is to calculate the *risk premium* R of the lottery, which is defined by (e.g. Kreps 1990, Varian 1992: 181)[24]

$$u\left(\mathcal{E}[y] - R\right) = \mathcal{E}\left[u(y)\right] \ . \tag{9.13}$$

The risk premium R is the amount of money that leaves a decision maker equally well-off, in terms of utility, between the two situations of (1) receiving for sure the expected pay-off from the lottery $\mathcal{E}[y]$ minus the risk premium R, and (2) playing the risky lottery with random pay-off y.[25] In general, if the utility function u characterizes a risk averse (risk neutral, risk loving) decision maker, the risk premium R is positive (zero, negative).

In the model employed here the risk premium of the lottery v depends on the level of biodiversity and is given by (see Appendix A9.2)

$$R(v) = \frac{\rho}{2}\,\sigma_s^2(v) \ . \tag{9.14}$$

The insurance value of biodiversity can now be defined based on the risk premium of the lottery v (Baumgärtner and Quaas 2005).

---

[24]By Equation (9.13), $\mathcal{E}[y] - R$ is the *certainty equivalent* of lottery v, as it yields the expected utility $\mathcal{E}\left[u(y)\right]$. According to Equations (9.3) and (9.5) $y \in Y$ with Y as an interval of $\mathbb{R}$, and according to Equation (9.9) u is continuous and strictly increasing, so that a risk premium R uniquely exists for every lottery v (Kreps 1990: 84).

[25]In the simple model employed here, the risk premium is equivalent to the so-called 'option price' of risk reduction, that is, the amount of money that a decision maker would be willing to pay for getting the expected pay-off from the lottery, $\mathcal{E}[y]$, for sure instead of playing the risky lottery with random pay-off y.

Definition 9.1
The *insurance value* V of biodiversity v is given by the change of the risk premium R of the lottery v due to a marginal change in the level of biodiversity v:

$$V(v) := -R'(v) . \tag{9.15}$$

Thus, the insurance value of biodiversity is the marginal value of biodiversity in its function to reduce the risk premium of the ecosystem manager's income risk from using ecosystem services under uncertainty. Being a marginal value, it depends on the existing level of biodiversity v. The minus sign in the defining Equation (9.15) serves to express biodiversity's ability to *reduce* the risk premium of the lottery v as a *positive* value. Applying Definition 9.1 to Equation (9.14), one obtains the following result for the insurance value of biodiversity in this model.

Proposition 9.1
The insurance value $V(v)$ of biodiversity is given by

$$V(v) = -\rho\,\sigma_s(v)\,\sigma_s{}'(v) > 0 . \tag{9.16}$$

From this equation it is apparent that the insurance value of biodiversity has an objective and a subjective dimension. The objective dimension is captured by the sensitivity of the standard deviation of ecosystem services to changes in biodiversity, σ_s and $\sigma_s{}'$; the subjective dimension is captured by the ecosystem manager's degree of risk aversion, ρ. The insurance value V increases

- with the degree ρ of the ecosystem manager's risk aversion and
- with the sensitivity of the standard deviation of ecosystem services to changes in biodiversity, σ_s and $|\sigma_s^{2\prime}|$.

As a function of biodiversity v, the insurance value $V(v)$ decreases (Figure 9.4): as biodiversity becomes more abundant (scarcer), its insurance value decreases (increases).

9.4.3 The Optimal Level of Biodiversity

In order to study how the ecosystem manager will make use of the insurance function of biodiversity, consider first the situation in which there is no financial insurance available. The ecosystem manager chooses a level of biodiversity v such as to maximize his expected utility (9.10):

$$\max_v \; U(v) . \tag{9.17}$$

With no financial insurance available, income y is given by Equation (9.5), such that the mean income μ_y and the standard deviation of income σ_y are given by Equations (9.6) and (9.7). The following proposition states the properties of the optimal solution to problem (9.17).

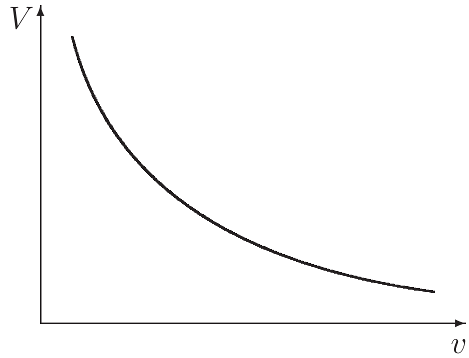

Figure 9.4 The insurance value V of biodiversity (Equation 9.16) as a function of existent biodiversity v.

Proposition 9.2

(i) The optimal level of biodiversity $v^\star$, which solves the ecosystem manager's optimization problem (9.17), is characterized by the necessary and sufficient condition

$$\mu_s'(v^\star) + V(v^\star) = C'(v^\star) \ . \tag{9.18}$$

(ii) The higher the ecosystem manager's degree of risk aversion ρ, the higher the optimal level of biodiversity $v^\star$:

$$\frac{dv^\star}{d\rho} > 0 \ . \tag{9.19}$$

Proof: see Appendix A9.3.

Condition (9.18) states that the optimal level of biodiversity $v^\star$ is chosen such that the marginal benefits of biodiversity equal its marginal costs. The marginal benefits here are composed of two additive components: the marginal gain in the mean level of ecosystem service and the insurance value $V(v^\star)$ of biodiversity. Hence, the insurance value of biodiversity is a value component in addition to the usual value arguments (such as direct or indirect use or non-use values, or existence values)[26] which hold in a world of certainty. It leads to choosing a higher level of biodiversity than without taking the insurance value into account.

The second part of the proposition states that the higher the degree of risk aversion ρ, the higher the optimal level of biodiversity $v^\star$. This is intuitively obvious, and confirms the idea that biodiversity is being used by a risk averse ecosystem manager as a form of natural insurance.

[26]See the detailed discussion on the economic value of biodiversity in Section 7.3.

9.4.4 The Effect of Financial Insurance

Consider now the situation in which there is financial insurance available. As an example, think again of the orchard farmer, who crucially depends on the pollination service provided by insects and who can manage his agro-ecosystem by choosing the level of biodiversity, e.g. by setting aside land for hedges and wetlands. As we have seen above, this farmer can manage his income risk from the random level of ecosystem service by choosing the level of biodiversity. On the other hand, the farmer may also have access to commercial crop yield insurance. Hence, his risk management now comprises two instruments. The ecosystem manager chooses a level of biodiversity v and a fraction of financial insurance coverage a such as to maximize his expected utility (9.10):

$$\max_{v,\,a} U(v, a) \ , \tag{9.20}$$

Income y is now given by

$$y = s - C(v) - a(s - \mu_s(v)) - \frac{\delta}{2}\, a^2\, \sigma_s^2(v) \ . \tag{9.21}$$

The first two components represent the benefits and costs of ecosystem management (Equation 9.5), the third component is the actuarially fair insurance premium/indemnification benefit (Equation 9.11) and the fourth component are the real costs of financial insurance (Equation 9.12). While the real costs of both ecosystem management and financial insurance (i.e. the second and fourth component) are certain, the benefits (i.e. the first and third component) are random. As a result, the mean and standard deviation of income are given by

$$\mu_y(v, a) \;=\; \mu_s(v) - C(v) - \frac{\delta}{2}\, a^2\, \sigma_s^2(v) \quad \text{and} \tag{9.22}$$

$$\sigma_y(v, a) \;=\; (1 - a)\, \sigma_s(v) \ . \tag{9.23}$$

Since the actuarially fair insurance premium/indemnification benefit corresponds to an expected payment of exactly zero, the mean income (Equation 9.22) is given by the mean benefits of ecosystem service minus the real costs of ecosystem management and financial insurance. The standard deviation of income (Equation 9.23) is given by the standard deviation of ecosystem service, reduced by a factor of $0 \le (1 - a) \le 1$. This should be compared to the case without financial insurance, where the standard deviation of income is given by the full standard deviation of ecosystem service (Equation 9.7). Equation (9.23) expresses the fact that the ecosystem manager can reduce the standard deviation of his income, besides by increasing the level of biodiversity v and thus lowering $\sigma_s(v)$, by increasing the fraction a of financial insurance coverage. In the extreme, with full coverage by financial insurance ($a = 1$) the standard deviation of income vanishes. With (9.22) and (9.23), the expected utility (9.10) is given by

$$\begin{aligned} U(v,a) &= \mu_y(v,a) - \frac{\rho}{2}\sigma_y^2(v,a) \\ &= \mu_s(v) - C(v) - \frac{\delta}{2}\,a^2\,\sigma_s^2(v) - \frac{\rho}{2}(1-a)^2\sigma_s^2(v)\ . \end{aligned} \tag{9.24}$$

The following proposition states the properties of the optimal solution to problem (9.20).

Proposition 9.3

(i) The optimal level of biodiversity $\hat{v}$ and the optimal fraction of financial insurance coverage $\hat{a}$, which solve the ecosystem manager's optimization problem (9.20), are characterized by the necessary and sufficient conditions

$$\mu_s'(\hat{v}) + \frac{\delta}{\rho+\delta}\,V(\hat{v}) = C'(\hat{v}) \qquad \textit{and} \tag{9.25}$$

$$\hat{a} = \frac{\rho}{\rho+\delta} \tag{9.26}$$

(ii) The higher the real costs of financial insurance, as measured by δ, the lower the optimal fraction $\hat{a}$ of coverage by financial insurance and the higher the optimal level $\hat{v}$ of biodiversity:

$$\frac{d\hat{a}}{d\delta} < 0 \qquad \textit{and} \qquad \frac{d\hat{v}}{d\delta} > 0\ . \tag{9.27}$$

(iii) A risk averse ecosystem manager chooses

- *full coverage by financial insurance ($\hat{a} = 1$) if $\delta = 0$,*
- *partial coverage by financial insurance ($0 < \hat{a} < 1$) if $0 < \delta < +\infty$, and*
- *no coverage by financial insurance ($\hat{a} \to 0$) if $\delta \to +\infty$.*

(iv) A risk averse ecosystem manager chooses $\hat{v} < v^\star$.

Proof: see Appendix A9.4.

The optimal allocation of biodiversity $\hat{v}$ and financial insurance coverage $\hat{a}$ is characterized by Conditions (9.25) and (9.26). Condition (9.25) states – similarly to Condition (9.18) in the absence of financial insurance – that the optimal level of biodiversity $\hat{v}$ is chosen such that the marginal benefits of biodiversity equal its marginal costs. The marginal benefits, again, are composed of two additive components: the marginal gain in the mean level of ecosystem service and the natural insurance value $V(\hat{v})$ of biodiversity, which is, however, not fully taken into account but only to a fraction $\delta/(\rho+\delta) < 1$. That is, biodiversity's natural insurance function is only partly taken into account when determining the optimal allocation.

The reason is that, of course, part of the income risk is now covered by financial insurance. Condition (9.26) specifies the optimal level of financial insurance coverage. It is obvious that $\hat{a}$ and the factor in front of $V(\hat{v})$ in Condition (9.25) add up to one. This means, biodiversity as the natural form of insurance and financial insurance together provide the optimal coverage of income risk.[27] Indeed, the two forms of insurance are substitutes: whatever part of the risk is not covered by biodiversity is covered by financial insurance. And what part of the risk is covered by financial insurance is determined by the real costs of financial insurance. Part (ii) of the proposition details this result: the higher the real costs of financial insurance, i.e. costs over and above the actuarially fair risk premium, the lower is the fraction of income risk covered by financial insurance and the higher is the fraction covered by the natural insurance, i.e. biodiversity.

Part (iii) of the proposition describes this in more detail. A risk averse ecosystem manager ($\rho > 0$) chooses full coverage by financial insurance ($\hat{a} = 1$) if it is available at actuarially fair conditions ($\delta = 0$); he chooses only partial coverage by financial insurance ($0 < \hat{a} < 1$) if financial insurance comes at additional costs over and above the actuarially fair risk premium ($0 < \delta < +\infty$); and he chooses no coverage by financial insurance ($\hat{a} \to 0$) if financial insurance becomes infinitely costly ($\delta \to +\infty$). These three cases imply, respectively, that the fraction of biodiversity's insurance value $V(\hat{v})$ which is taken into account according to condition (9.25), which is also the fraction of income risk covered by the natural insurance of biodiversity, is zero if financial insurance is available at actuarially fair conditions; it is in between zero and one if financial insurance is available at actuarially unfair conditions; and it goes to one for infinitely unfair financial insurance.

Part (iv) of the proposition states that in any case, a risk averse ecosystem manager chooses a lower level of biodiversity if financial insurance is available compared to a situation where no financial insurance is available: $\hat{v} < v^{\star}$. That is, financial insurance crowds out biodiversity as the natural from of insurance.

[27]Note that this does *not* necessarily mean that in the optimal allocation there is no more income risk, i.e. $\sigma_y^2(\hat{v}, \hat{a}) = 0$. It only means that the overall amount of income variance that the decision maker wishes to avoid in the optimum is covered by both natural and financial insurance. This may still leave the decision maker with some positive income risk in the optimum, i.e. $\sigma_y^2(\hat{v}, \hat{a}) > 0$.

9.5 Discussion

Although the results have been derived from a very simple and specific model, they are robust to a fair amount of generalization. For instance, while the choice of the preference representation (9.9) served to obtain simple closed-form solutions, all results thus obtained are qualitatively robust to generalizations to expected utility functions of the type $U(\mu_y, \sigma_y^2)$ with $\partial U/\partial \mu_y > 0$ and $\partial U/\partial \sigma_y^2 < 0$. Also, while the specific form of financial insurance contract assumed here (Section 9.3.3) served to obtain simple closed-form solutions, all results thus obtained are qualitatively robust to generalizations to more general financial insurance contracts with an actuarially fair insurance premium plus a transaction costs/profit mark-up on top (Quaas and Baumgärtner 2006). And while I have assumed for simplicity that the level of biodiversity is the only determinant of the statistical distribution of the ecosystem service (Equations 9.1, 9.2), one could easily generalize the analysis so that the stochastic production of the ecosystem service depends also on inputs other than biodiversity, say labor, capital, fertilizer or chemical pest control. This could be formalized with the help of a Just-Pope-production function (Just and Pope 1978, 1979), which is well suited for mean-variance-analysis of stochastic production, and would not qualitatively alter the basic results about the role of biodiversity for income risk.

Of particular importance are Assumptions (9.1), (9.2) and (9.3) about the benefits and costs of biodiversity. While Assumptions (9.1) and (9.2) represent the best available ecological knowledge and describe a relevant problem, it is an interesting question whether these assumptions are actually necessary in order to arrive at the main result (i.e. biodiversity's insurance value) or whether this result holds under more general conditions. It turns out that the crucial assumption is $\sigma_s' < 0$, while $\mu_s' > 0$ is not necessary. If biodiversity did depress the mean level of ecosystem services ($\mu_s' < 0$) then this could be considered as costs of biodiversity and could be included in the function $C(v)$. This assumption would therefore not lead to a different result. If, however, biodiversity did increase the variance of ecosystem services ($\sigma_s' > 0$) then it would obviously not have any insurance value. Clearly, this would fundamentally alter the main results of the paper. As for the assumption on second derivatives ($\mu_s' \leq 0$, $\sigma_s' \geq 0$, $C' \geq 0$), their role is mainly technical, making sure that second order conditions are fulfilled and that one has an interior solution. Without these assumptions, the main results would not change fundamentally, but would require a more elaborate formulation and proof of results.

So, the crucial assumptions which ultimately limit the generality of results are the following:

- The ecosystem manager is risk averse and maximizes his expected utility from an uncertain income which is determined by the random level of some ecosystem service.

- The level of biodiversity determines the probability distribution of the ecosystem service and, thus, of income. Taking into account the (direct or opportunity) costs of biodiversity, there is a positive correlation between expected income and standard deviation of income in the relevant range of feasible income distributions.
- A financial insurance contract specifies only the state dependent redemption payment and the corresponding risk premium. In particular, it is not explicitly contingent on the particular level of biodiversity chosen by the ecosystem manager.[28]
- Both insurant and insurance company have the same ex ante knowledge about the probability distribution of ecosystem services. Both can observe ex post the actual state of nature.

While these assumptions limit the generality of the results obtained here, they describe – in a very stylized way – a realistic scenario of managing stochastic ecosystems under uncertainty for the ecosystem services they provide. Hence, this analysis yields relevant insights into the issue.

9.6 Conclusion

I have presented a conceptual ecological-economic model that combines (i) ecological results about the relationships between biodiversity, ecosystem functioning, and the provision of ecosystem services with (ii) economic methods to study decision-making under uncertainty. In this framework I have (1) determined the insurance value of biodiversity, (2) studied the optimal allocation of funds in the trade-off between investing into biodiversity protection and the

[28]This gives rise to what is known in the insurance economics literature as 'moral hazard' (Kreps 1990). As long as the behavior of the economic agent (here: the level of biodiversity chosen be the ecosystem manager) cannot be observed by the insurance company, but only the resulting outcome can be observed (here: the provision of some ecosystem service), the existence of insurance will induce the insurant to choose a riskier behavior than if insurance was not available. Moral hazard is a problem for many insurance markets, e.g. health insurance or car insurance, and has been identified as a major reason for the absence of private insurance markets for most agricultural risks (Chambers 1989). Because of the moral hazard problem, most insurance contracts intentionally do not allow for full coverage at actuarially fair premiums, but contain deductibles or upper limits in either the degree of coverage or the amount to be insured. Other insurance policies try to include a specification of the insurant's behavior (or observable proxies thereof) into the contract. These mechanisms serve to diminish the moral hazard problem, yet they cannot eliminate it completely.

purchase of financial insurance, and (3) analyzed the effect of different institutional settings in the market for financial insurance on biodiversity protection. The focus was on how to model the ecology-economy-interface. Relevant economic and policy questions that arise from this view on biodiversity – e.g. the public good character of the problem, the dynamics of the problem or implications for environmental and development policies – are discussed in more detail elsewhere (Baumgärtner and Quaas 2005, Quaas and Baumgärtner 2006, Quaas et al. 2004).

The conclusion from this analysis is that biodiversity can be interpreted as a form of natural insurance for risk averse ecosystem managers against the over- or under-provision with ecosystem services, such as biomass production, control of water run-off, pollination, control of pests and diseases, nitrogen fixation, soil regeneration etc. Thus, biodiversity has an insurance value, which is a value component in addition to the usual value arguments (such as direct or indirect use or non-use values, or existence values) holding in a world of certainty. This insurance value should be taken into account when deciding upon how much to invest into biodiversity protection. It leads to choosing a higher level of biodiversity than without taking the insurance value into account, with a higher degree of risk aversion leading to a higher optimal level of biodiversity. As far as the insurance function is concerned, biodiversity and financial insurance against income risk, e.g. crop yield insurance, may be seen as substitutes. If financial insurance is available, a risk averse ecosystem manager, say, a farmer, will partially or fully substitute biodiversity's insurance function by financial insurance, with the extent of substitution depending on the costs of financial insurance. Hence, the availability, and exact institutional design, of financial insurance influence the level of biodiversity protection.

Appendix

A9.1 Expected Utility Function (9.10)

With

$$f(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}}\, e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} \tag{A9.1}$$

as the probability density function of the normal distribution of income y with mean μ_y and variance σ_y^2, the von (Neumann-Morgenstern) expected utility from the (Bernoulli) utility function (9.9) is

$$\tilde{U} = \mathcal{E}[u(y)] = -\int e^{-\rho\, y} f(y) dy = -e^{-\rho\left[\mu_y - \frac{\rho}{2}\sigma_y^2\right]} \,. \tag{A9.2}$$

Using a simple monotonic transformation of $\tilde{U}$, one obtains the expected utility function U (Equation 9.10).

A9.2 Risk Premium (9.14)

The risk premium R has been defined in Equation (9.13) as

$$u\left(\mathcal{E}[y] - R\right) = \mathcal{E}\left[u(y)\right] \ . \tag{A9.3}$$

With the Bernoulli utility function (9.9) and $\mathcal{E}[y] = \mu_y$ the left hand side of this equation is given by

$$u\left(\mathcal{E}[y] - R\right) = -e^{-\rho\,[\mu_y - R]} \ , \tag{A9.4}$$

and the right hand side is given by Equation (A9.2). Hence, we have

$$-e^{-\rho\,[\mu_y - R]} = -e^{-\rho\left[\mu_y - \frac{\rho}{2}\,\sigma_y^2\right]} \ . \tag{A9.5}$$

Rearranging, and observing that $\sigma_y^2 = \sigma_s^2$ (Equation 9.7), yields the result stated in Equation (9.14).

A9.3 Proof of Proposition 9.2

ad (i). In Problem (9.17), the objective function to be maximized over v is

$$U(v) = \mu_s(v) - C(v) - \frac{\rho}{2}\sigma_s^2(v) \ , \tag{A9.6}$$

such that the first order condition for a solution $v^\star$ is

$$\mu_s'(v^\star) - C'(v^\star) - \rho\,\sigma_s(v^\star)\,\sigma_s'(v^\star) = 0 \ . \tag{A9.7}$$

Observing that $-\rho\,\sigma_s(v^\star)\,\sigma_s'(v^\star) = V(v^\star)$ (Equation 9.16) yields Equation (9.18). With Assumptions (9.1), (9.2) and (9.3) about the curvature of these functions, the second order condition for a maximum,

$$\mu_s''(v^\star) - C''(v^\star) - \rho\left(\sigma_s'(v^\star)\right)^2 - \rho\,\sigma_s(v^\star)\,\sigma_s''(v^\star) < 0 \ , \tag{A9.8}$$

is satisfied, such that the necessary first order condition is also sufficient.

ad (ii). The total derivative of first order condition (9.18) with respect to ρ is

$$\mu_s''\,\frac{dv^\star}{d\rho} - C''\,\frac{dv^\star}{d\rho} + V'\,\frac{dv^\star}{d\rho} - \sigma_s\sigma_s' = 0 \ . \tag{A9.9}$$

This can be rearranged into

$$\frac{dv^\star}{d\rho} = \frac{\sigma_s\,\sigma_s'}{\mu_s'' - C'' + V'} > 0 \ , \tag{A9.10}$$

which is strictly positive due to Assumptions (9.1), (9.2), (9.3) and $V' < 0$ (Proposition 9.1).

A9.4 Proof of Proposition 9.3

ad (i). In Problem (9.20), the objective function to be maximized over v and a is

$$U(v,a) = \mu_s(v) - C(v) - \frac{\delta}{2}\, a^2\, \sigma_s^2(v) - \frac{\rho}{2}(1-a)^2\, \sigma_s^2(v)\ , \tag{A9.11}$$

such that the first order conditions for a solution $(\hat{v},\hat{a})$ are

$$\begin{aligned} U_v(\hat{v},\hat{a}) &= \mu_s'(\hat{v}) - C'(\hat{v}) - \delta \hat{a}^2 \sigma_s(\hat{v})\sigma_s'(\hat{v}) - \rho(1-\hat{a})^2 \sigma_s(\hat{v})\sigma_s'(\hat{v}) = 0, & \text{(A9.12)}\\ U_a(\hat{v},\hat{a}) &= -\delta \hat{a} \sigma_s^2(\hat{v}) + \rho(1-\hat{a})\sigma_s^2(\hat{v}) = 0\ . & \text{(A9.13)} \end{aligned}$$

As $\sigma_s^2(v) > 0$ for all v, Condition (A9.13) can be solved to yield

$$\hat{a} = \frac{\rho}{\rho+\delta}\ , \tag{A9.14}$$

which is the result stated in the proposition (Equation 9.26). This can be inserted into Condition (A9.12), which yields, after rearranging,

$$\mu_s'(\hat{v}) + \frac{\delta}{\rho+\delta}\,(-\rho\, \sigma_s(\hat{v})\, \sigma_s'(\hat{v})) = C'(\hat{v})\ . \tag{A9.15}$$

Observing that $-\rho\, \sigma_s(\hat{v})\, \sigma_s'(\hat{v}) = V(\hat{v})$ (Equation 9.16) yields Equation (9.25). As for the second order condition, note that

$$\begin{aligned} U_{vv}(\hat{v},\hat{a}) &= \mu_s''(\hat{v}) - C''(\hat{v}) - \delta\, \hat{a}^2\, (\sigma_s'(\hat{v}))^2 - \delta\, \hat{a}\, \sigma_s(\hat{v})\, \sigma_s''(\hat{v}) \\ &\quad -\rho\,(1-\hat{a})^2\, (\sigma_s'(\hat{v}))^2 - \rho\,(1-\hat{a})^2\, \sigma_s(\hat{v})\, \sigma_s''(\hat{v}) < 0\ , & \text{(A9.16)}\\ U_{aa}(\hat{v},\hat{a}) &= -(\delta+\rho)\sigma_s^2(\hat{v}) < 0\ , & \text{(A9.17)}\\ U_{va}(\hat{v},\hat{a}) &= -2\delta\hat{a}\, \sigma_s(\hat{v})\, \sigma_s'(\hat{v}) + 2\rho(1-\hat{a})\, \sigma_s(\hat{v})\, \sigma_s'(\hat{v}) = 0\ , & \text{(A9.18)} \end{aligned}$$

where the last equality follows from using first order condition (A9.14). Hence, $U_{vv}U_{aa} - U_{va}^2 > 0$, so that the second order condition for a maximum is satisfied and the necessary first order conditions are also sufficient.

ad (ii). The total derivative of first order condition (9.25) with respect to δ is

$$\mu_s''\, \frac{d\hat{v}}{d\delta} + \frac{\rho}{(\rho+\delta)^2}\, V + \frac{\delta}{\rho+\delta}\, V'\, \frac{d\hat{v}}{d\delta} - C''\, \frac{d\hat{v}}{d\delta} = 0\ , \tag{A9.19}$$

which can be rearranged into

$$\frac{d\hat{v}}{d\delta} = -\frac{\frac{\rho}{(\rho+\delta)^2} V}{\mu_s'' + \frac{\delta}{\rho+\delta} V' - C''} > 0\ , \tag{A9.20}$$

which is strictly positive due to Assumptions (9.1), (9.3) and $V' < 0$ (Proposition 9.1). The result about $d\hat{a}/d\delta$ follows immediately from Condition (9.26).

Part (iii) of the proposition follows immediately from Condition (9.26).

ad (iv). Compare Conditions (9.18) and (9.25) for $v^\star$ and $\hat{v}$ respectively in a slightly rearranged version:

$$\mu_s'(v^\star) - C'(v^\star) \quad = \quad -V(v^\star) \; , \tag{A9.21}$$

$$\mu_s'(\hat{v}) \; - C'(\hat{v}) \quad = \quad \frac{\delta}{\rho + \delta} \left(-V(v^\star) \right) \; . \tag{A9.22}$$

From Assumptions (9.1) and (9.3) it follows that $\mu_s'(v) - C'(v)$ is a decreasing function of v, while it follows from Proposition 9.1 that $-V(v)$ is an increasing function of v, so that $v^\star$ and $\hat{v}$ are determined by the intersection of the decreasing curve representing the left-hand-side and the increasing curve representing the right-hand side of Conditions (A9.21) and (A9.22) respectively. The difference between these two conditions is that for every v the function on the right-hand side of Condition (A9.22) yields smaller values than the one in Condition (A9.21), as $0 < \delta/(\rho + \delta) < 1$, so that the intersection determining the optimal v in Condition (A9.22) is further to the left than the one in Condition (A9.21), i.e. $\hat{v} < v^\star$.

10. Insurance and Sustainability through Ecosystem Management*

with Christian Becker, Karin Frank, Birgit Müller and Martin Quaas

10.1 Introduction

There is a widely held belief that individual myopic optimization is at odds with long-term sustainability of an ecological-economic system. In this paper, we want to take a fresh look at this position. We show that for typical ecosystems and under plausible and standard assumptions about individual decision making, myopic optimization may lead to sustainable outcomes. In particular, in order to explain the sustainable use of ecosystems, it is not necessary to assume preferences for sustainability – or any special concern for the distant future – on the part of the decision maker; it suffices to assume that a myopic decision maker is sufficiently risk averse.

The ecological-economic system under study here is grazing in semi-arid rangelands. Semi-arid regions cover one third of the Earth's land surface. They are characterized by low and highly variable precipitation. Their utilization in livestock farming provides the livelihood for a large part of the local populations. Yet, over-utilization and non-adapted grazing strategies lead to environmental problems such as desertification.

Grazing in semi-arid rangelands is a prime object of study for ecological economics, as the ecological and economic systems are tightly coupled (e.g. Beukes et al. 2002, Heady 1999, Janssen et al. 2004, Perrings 1997, Perrings and Walker 1997, 2004, Westoby et al. 1989). The grass biomass is directly used as forage for livestock, which is the main source of income; and the grazing pressure from livestock farming directly influences the ecological dynamics. The crucial link is the grazing management.

*Forthcoming in *Ecological Economics* as 'Uncertainty and sustainability in the management of rangelands'.

The ecological dynamics, and thus, a farmer's income, essentially depend on the low and highly variable rainfall. The choice of a properly adapted grazing management strategy is crucial in two respects: first, to maintain the rangeland system as an income base, that is, to prevent desertification; and second, to smooth out income fluctuations, in particular, to avoid high losses in the face of droughts.

Assuming that the farmer is non-satiated in income and risk averse, we analyze the choice of a grazing management strategy from two perspectives. In a first step we determine a myopic farmer's optimal grazing management strategy. We show that a risk averse farmer chooses a strategy in order to obtain 'insurance' from the ecosystem (Baumgärtner and Quaas 2005). That is, the optimal strategy reduces income variability, but yields less mean income than possible. In a second step we analyze the long-term ecological and economic impact of different strategies. We conclude that the more risk averse a myopic farmer is, the more conservative is his optimal grazing management strategy. If he is sufficiently risk averse, the optimal strategy is conservative enough to be sustainable.

Following the literature on grazing management under uncertainty, we analyze the choice of a *stocking rate* of livestock, as this is the most important aspect of rangeland management (e.g. Hein and Weikard 2004, Karp and Pope 1984, McArthur and Dillon 1971, Perrings 1997, Rodriguez and Taylor 1988, Torell et al. 1991, Westoby et al. 1989). The innovative analytical approach taken here is to consider the choice of a *grazing management strategy*, which is a *rule* about the stocking rate to apply in any given year depending on the rainfall in that year. This is inspired by empirical observations in Southern Africa. Rule-based grazing management has the twofold advantage that a farmer has to make a choice (concerning the rule) only once, and yet, keeps a certain flexibility and scope for adaptive management (concerning the stocking rate). The flexibility thus obtained is the decisive advantage of choosing a constant rule over choosing a constant stocking rate.

The paper is organized as follows. In Section 10.2, we discuss grazing management in semi-arid rangelands in more detail and describe one particular 'good practice'-example: the Gamis Farm, Namibia. In Section 10.3, we develop a dynamic and stochastic ecological-economic model, which captures the essential aspects and principles of grazing management in semi-arid rangelands, and features the key aspect of the Gamis-strategy. Our results are presented in Section 10.4, with all derivations and proofs given in the Appendix. Section 10.5 concludes.

10.2 Grazing Management in Semi-Arid Rangelands: The Gamis Farm, Namibia

The dynamics of ecosystems in semi-arid regions are essentially driven by low and highly variable precipitation (Behnke et al. 1993, Sullivan and Rhode 2002, Westoby et al. 1989).[1] Sustainable economic use of these ecosystems requires an adequate adaption to this environment. The only sensible economic use, which is indeed predominant (Mendelsohn et al. 2002), is by extensive livestock farming. However, over-utilization and inadequate management lead to pasture degradation and desertification. Rangeland scientists have proposed different types of grazing management strategies in order to solve these problems. A low constant stocking rate was recommended by Lamprey (1983) and Dean and Mac Donald (1994), who assumed that grazing pressure is the main driving force for vegetation change and that rangeland systems reach an equilibrium state. Other authors considered the highly variable rainfall to be the major driving force and claimed that grazing has only marginal influence on vegetation dynamics (Behnke et al. 1993, Scoones 1994, Sandford 1994, Westoby et al. 1989). They recommend an 'opportunistic' strategy which matches the stocking rate with the available forage in every year. Thus, the stocking rate should be high in years with sufficient rainfall, and low when there is little forage in dry years (Beukes et al. 2002: 238). Recent studies have shown that both grazing and variable rainfall are essential for the vegetation dynamics on different temporal and spatial scales (Cowling 2000, Briske et al. 2003, Fuhlendorf and Engle 2001, Illius and O'Connor 1999, 2000, Vetter 2005).

One example of a sophisticated and particularly successful grazing management system has been employed for forty years at the Gamis Farm, Namibia (Müller et al. forthcoming, Stephan et al. 1996, 1998a, 1998b). The Gamis Farm is located 250 km southwest of Windhoek in Namibia (24°05′S 16°30′E) close to the Naukluft mountains at an altitude of 1,250 m. The climate of this arid region is characterized by low mean annual precipitation (177 mm/y) and high variability (variation coefficient: 56 %). The vegetation type is dwarf shrub savanna (Giess 1998); the grass layer is dominated by the perennial grasses *Stipagrostis uniplumis*, *Eragrostis nindensis* and *Triraphis ramosissima* (Maurer 1995).

Karakul sheep (race Swakara) are bred on an area of 30,000 hectares. The primary source of revenue is from the sale of lamb pelts. Additionally, the wool of the sheep is sold. In good years, up to 3,000 sheep are kept on the farm. An adaptive grazing management strategy is employed to cope with the variability in forage. The basis of the strategy is a rotational grazing scheme:

---

[1] Another important driver of ecological dynamics in semi-arid rangelands is the stochastic occurrence of fire (Janssen et al. 2004, Perrings and Walker 1997, 2004). In our case, fire plays only a minor role, but the stochasticity of rainfall is crucial (Müller et al. forthcoming).

the pasture land is divided into 98 paddocks, each of which is grazed for a short period (about 14 days) until the palatable biomass on that paddock is used up completely, and then is rested for a minimum of two months. This system puts high pressure on the vegetation for a short time to prevent selective grazing (Batabyal and Beladi 2002, Batabyal et al. 2001, Heady 1999). While such a rotational grazing scheme is fairly standard throughout semi-arid regions, the farmer on the Gamis Farm has introduced an additional resting: in years with sufficient precipitation one third of the paddocks are given a rest during the growth period (September – May). In years with insufficient rainfall this rest period is reduced or completely omitted. Once a year, at the end of the rainy season (April), the farmer determines – based on actual rainfall and available forage – how many paddocks will be rested and, thus, how many lambs can be reared. This strategy is a particular example of what has been called 'rotational resting' (Heady 1970, 1999, Stuth and Maraschin 2000, Quirk 2002).

The grazing management system employed at the Gamis Farm has been successful over decades, both in ecological and economic terms. It, therefore, represents a model for commercial farming in semi-arid rangelands.

10.3 The Model

Our analysis is based on an integrated dynamic and stochastic ecological-economic model, which captures essential aspects and principles of grazing management in semi-arid regions. It represents a dynamic ecosystem, which is driven by stochastic precipitation, and a risk averse farmer, who rationally chooses a grazing management strategy under uncertainty.

10.3.1 Precipitation

Uncertainty is introduced into the model by the stochasticity of rainfall, which is assumed to be an independent and identically distributed (iid) random variable. For semi-arid areas, a log-normal distribution of rainfall $r(t)$ is an adequate description (Sandford 1982).[2] The log-normal distribution, with probability density function $f(r)$ (Equation A10.1), is determined by the mean μ_r and standard deviation σ_r of precipitation. Here, we measure precipitation in terms of 'ecologically effective rain events', i.e. the number of rain events during rainy season with a sufficient amount of rainfall to be ecologically productive (Müller et al. forthcoming).

[2]While the distribution of rainfall $r(t)$ is exogenous, all other random variables in the model follow an induced distribution.

10.3.2 Grazing Management Strategies

The farm is divided into a number $I \in \mathbb{N}$ of identical paddocks, numbered by $i \in \{1, \ldots, I\}$. In modeling grazing management strategies, we focus on the aspect of additional resting during the growth period, which is the innovative element in the Gamis grazing system. That is, we analyze rotational resting of paddocks from year to year, but do not explicitly consider rotational grazing during the year (cf. Section 10.2). The strategy is applied in each year, after observing the actual rainfall at the end of the rainy season. Its key feature is that in dry years all paddocks are used, while in years with sufficient rainfall a pre-specified fraction of paddocks is rested. Whether resting takes place, and to what extent, are the defining elements of what we call the farmer's grazing management strategy:

Definition 10.1
A *grazing management strategy* $(\alpha, \underline{r})$ is a rule of how many paddocks are not grazed in a particular year given the actual rainfall in that year, where $\alpha \in [0, 1]$ is the fraction of paddocks rested if rainfall exceeds the threshold value $\underline{r} \in [0, \infty)$.[3]

Thus, when deciding on the grazing management strategy, the farmer decides on two variables: the rain threshold $\underline{r}$ and the fraction α of rested paddocks. While the rule is constant (i.e. $\alpha =$ const., $\underline{r} =$ const.) its application may yield a different stocking with livestock in any given year depending on actual rainfall in that year.

In the resource economics literature, this type of strategy is called 'proportional threshold harvesting' (Lande et al. 2003). This is a form of adaptive management: the (constant) rule adapts the fraction of fallow paddocks and the number of livestock kept on the farm as actual rainfall changes. Note that the 'opportunistic' strategy (e.g. Beukes et al. 2002: 238) is the special case without resting, i.e. $\alpha = 0$.

10.3.3 Ecosystem Dynamics

Both the stochastic rainfall and grazing pressure are major determinants of the ecological dynamics. Following Stephan et al. (1998a), we consider two quantities to describe the state of the vegetation in each paddock i at time t: the green biomass $G^i(t)$ and the reserve biomass $R^i(t)$ of a representative grass species,[4] both of which are random variables, since they depend on the random

---

[3]We assume that the number I of paddocks is so large that we can treat α as a real number.

[4]We assume that selective grazing is completely prevented, i.e. there is no competitive disadvantage for more palatable grasses (see e.g. Beukes et al. 2002). Hence, we consider a single, representative species of grass.

variable rainfall. The green biomass captures all photosynthetic ('green') parts of the plants, while the reserve biomass captures the non-photosynthetic reserve organs ('brown' parts) of the plants below or above ground (Noy-Meir 1982). The green biomass grows during the growing season in each year and dies almost completely in the course of the dry season. The amount $G^i(t)$ of green biomass available on paddock i in year t after the end of the growing season depends on rainfall $r(t)$ in the current year, on the reserve biomass $R^i(t)$ on that paddock, and on a growth parameter w_G:

$$G^i(t) = w_G \cdot r(t) \cdot R^i(t). \tag{10.1}$$

As the green biomass in the current year does not directly depend on the green biomass in past years, it is a flow variable rather than a stock.

In contrast, the reserve biomass $R^i(t)$ on paddock i in year t is a stock variable. That is, the reserve biomass parts of the grass survive several years ('perennial grass'). Thereby, the dynamics of the vegetation is not only influenced by the current precipitation, but also depends on the precipitation of preceding years (O'Connor and Everson 1998). Growth of the reserve biomass from the current year to the next one is

$$R^i(t+1)-R^i(t) = -d\cdot R^i(t)\cdot\left(1+\frac{R^i(t)}{K}\right)+w_R\cdot(1-c\cdot x^i(t))\cdot G^i(t)\cdot\left(1-\frac{R^i(t)}{K}\right), \tag{10.2}$$

where w_R is a growth parameter and d is a constant death rate of the reserve biomass, which we assume to be sufficiently small, i.e. $d < w_R\, w_G\, \mu_r$. A density dependence of reserve biomass growth is captured by the factors containing the capacity limits K: The higher the reserve biomass on paddock i, the slower it grows. The status variable x^i captures the impact of grazing on the reserve biomass of paddock i. If paddock i is grazed in year t, we set $x^i(t) = 1$, if it is rested, we set $x^i(t) = 0$. The parameter c (with $0 \le c \le 1$) describes the amount by which reserve biomass growth is reduced due to grazing pressure. For simplicity, we assume that the initial ($t = 1$) stock of reserve biomass of all paddocks is equal,

$$R^i(1) = R \quad \text{for all } i = 1, \ldots, I. \tag{10.3}$$

10.3.4 Livestock and Income

As for the dynamics of livestock, the herd size $S(t)$, that can be kept on the farm at time t, is limited by total available forage. We normalize the units of green biomass in such a way that one unit of green biomass equals the need of one livestock unit per year. Thus, total available green biomass on the farm, $\sum_{i=1}^{I} G^i(t)$, determines the 'carrying capacity', i.e. the maximum number of livestock that can be held on the farm in the period under consideration.[5] In general, the farmer will not stock up to this carrying capacity in every year. Rather, the herd size kept on the farm in period t is given by

$$S(t) = \sum_{i=1}^{I} x^i(t) \cdot G^i(t) \ . \tag{10.4}$$

That is, the herd size in year t is determined by the total green biomass of the paddocks used for grazing (i.e., not rested) in that year. For the sake of the analysis, we assume that the farmer annually rents his livestock on a perfect rental market for livestock.[6] This allows the farmer to exactly adapt the actual herd size to the available forage and to his chosen grazing management strategy.[7]

The herd size $S(t)$ kept on the farm in year t determines the farmer's income $y(t)$. We assume that the quantity of marketable products from livestock, e.g. lamb furs and wool, is proportional to the herd size. Normalizing product units in an appropriate way, the numerical value of output equals livestock $S(t)$. The farmer sells his products on a world market at a given price and takes the annual rental rate of livestock as given. The difference between the two is the net revenue per livestock unit, p. Assuming that farming is profitable, i.e. $p > 0$, the farmer's income $y(t)$ is

$$y(t) = p \cdot S(t). \tag{10.5}$$

Since the herd size $S(t)$ is a random variable, income $y(t)$ is a random variable, too.[8] In order to simplify the notation in the subsequent analysis, we normalize

$$p \equiv (w_G \cdot I \cdot R)^{-1}. \tag{10.6}$$

This means, from now on we measure net revenue per livestock unit in units of total forage per unit of precipitation. As a result, income is measured in units of precipitation.

---

[5]In contrast to the capacity limit K of reserve biomass, the carrying capacity of livestock is not a constant, but it depends on rainfall and the stock of reserve biomass (cf. Equation 10.1), and, therefore, will change over time.

[6]If the farmer owns a constant herd of size S_0, he would rent a number $S(t) - S_0$ if $S(t) > S_0$ or rent out a number $S_0 - S(t)$ if $S(t) < S_0$. Without loss of generality, we set $S_0 = 0$.

[7]Hence, the herd size $S(t)$ does not follow its own dynamics, but it is determined by precipitation and the chosen strategy.

[8]In our analysis, we neglect uncertainty of prices. Including a price stochasticity uncorrelated to rainfall would not alter our results. Including a price stochasticity with a negative correlation to rainfall would most likely reinforce our central result that a risk averse farmer chooses a conservative grazing management strategy, since high stocking rates in good rainy years become less valuable (as indicated by Hein and Weikard 2004).

For the subsequent analysis of a myopic farmer's decision, first and second year income are of particular interest. Given the actual rainfall $r(1)$ in the first grazing period, the initial reserve biomass (Equation 10.3) and a grazing management rule $(\alpha, \underline{r})$, the herd size $S(1)$ is determined by Equation (10.4). Inserting Equation (10.1) and using Assumption (10.3), as well as normalization (10.6), the farmer's income $y(1)$ in the first grazing period is given by Equation (10.5) as

$$y(1) = \frac{1}{I}\sum_{i=1}^{I} x^i(1)\cdot r(1) = \left\{ \begin{array}{cl} r(1) & \text{if} \quad r(1) \le \underline{r} \\ (1-\alpha)\cdot r(1) & \text{if} \quad r(1) > \underline{r} \end{array} \right. . \tag{10.7}$$

Given the probability density distribution $f(r)$ of rainfall, the mean $\mu_{y(1)}(\alpha, \underline{r})$ and the standard deviation $\sigma_{y(1)}(\alpha, \underline{r})$ of the first period's income are (see Appendix A10.1)

$$\mu_{y(1)}(\alpha, \underline{r}) = \mu_r - \alpha \int_{\underline{r}}^{\infty} r\, f(r)\, dr \tag{10.8}$$

$$\sigma_{y(1)}(\alpha, \underline{r}) = \sqrt{\sigma_r^2 + 2\,\alpha\,\mu_r \int_{\underline{r}}^{\infty} r\, f(r)\, dr - \alpha^2 \left[\int_{\underline{r}}^{\infty} r\, f(r)\, dr\right]^2 - \alpha\,(2-\alpha) \int_{\underline{r}}^{\infty} r^2\, f(r)\, dr}, \tag{10.9}$$

where μ_r and σ_r are the mean and the standard deviation of rainfall.

The model implies that resting in the first period has a positive impact on reserve biomass and, thus, on future income. In particular, if the farmer applies a grazing management strategy $(\alpha, \underline{r})$ with $\alpha > 0$ and $\underline{r} < \infty$, rather than full stocking, he can gain an extra income in the second year. Given the actual rainfall $r(1)$ in the first year, the additional reserve biomass in the second year is (cf. Equations 10.1, 10.2 and 10.3)

$$\Delta R = w_R \cdot w_G \cdot I \cdot R \cdot \left(1 - \frac{R}{K}\right) \cdot r(1) \cdot \left\{ \begin{array}{cl} 0 & \text{if} \quad r(1) \le \underline{r} \\ \alpha & \text{if} \quad r(1) > \underline{r} \end{array} \right. . \tag{10.10}$$

This additional reserve biomass gives rise to extra green biomass growth, and, hence, to additional income in the second year (cf. Equations 10.1, 10.4, 10.5 and 10.10):

$$\Delta y(2) = w_G \cdot r(2) \cdot \left\{ \begin{array}{cl} 1 & \text{if} \quad r(2) \le \underline{r} \\ 1-\alpha & \text{if} \quad r(2) > \underline{r} \end{array} \right\} \cdot w_R \cdot \left(1 - \frac{R}{K}\right) \cdot r(1) \cdot \left\{ \begin{array}{cl} 0 & \text{if} \quad r(1) \le \underline{r} \\ \alpha & \text{if} \quad r(1) > \underline{r} \end{array} \right. . \tag{10.11}$$

This means, the reserve biomass can be used as a buffer: by applying a grazing strategy with resting, the farmer can shift income to the next year. For a risk averse farmer, this extra income is particularly valuable if the second year is a dry year.

10.3.5 Farmer's Choice of Grazing Management Strategy

We assume that the farmer's utility only depends on income y, and that he is a non-satiated and risk averse expected utility maximizer. Let

$$U = \sum_{t=1}^{\infty} \frac{\mathcal{E}_t \, u(y(t))}{(1+\delta)^{t-1}} \tag{10.12}$$

be his von Neumann-Morgenstern intertemporal expected utility function, where δ is the discount rate, the Bernoulli utility function $u(\cdot)$ is a strictly concave function of income y, and $\mathcal{E}_t$ is the expectancy operator at time t. In particular, we employ a utility function with constant relative risk aversion,

$$u(y) = \frac{y^{1-\rho} - 1}{1-\rho}, \tag{10.13}$$

where $\rho > 0$ is the constant parameter which measures the degree of relative risk aversion (Gollier 2001).

The farmer will choose the grazing management strategy which maximizes his von Neumann-Morgenstern intertemporal expected utility function (10.12). The basic idea is to regard the choice of a grazing management strategy as the choice of a 'lottery' (Baumgärtner and Quaas 2005). Each possible lottery is characterized by the probability distribution of pay-off, where the pay-off is given by the farmer's income. Given the ecological dynamics, both the mean income and the standard deviation solely depend on the grazing management strategy applied. Thus, choosing a grazing management strategy implies choosing a particular distribution of income.

We assume that the farmer initially, i.e. at $t = 0$ prior to the first grazing period, chooses a grazing management strategy $(\alpha, \underline{r})$, which is then applied in all subsequent years. When choosing the strategy, the farmer does not know which amount of rainfall will actually occur, but he knows the probability distribution of rainfall. As a result, he knows the probability distribution of his income for any possible grazing management strategy. A far-sighted farmer would choose the grazing management strategy that maximizes his intertemporal utility (10.12), taking into account the effect of the strategy on the ecosystem dynamics, as given by Equations (10.1) and (10.2). In particular, he would account for the effect that resting improves the reserve biomass in the long run, compared to a strategy with full stocking. However, our aim is to show that a sufficiently risk averse farmer will choose a conservative strategy, even if he does not consider the long-term benefits. For this sake, we assume that the farmer is myopic in the following sense (Kurz 1987):

Definition 10.2
A *myopic farmer* neglects the long-term effects of his grazing management strategy on the ecosystem: (i) He assumes that reserve biomass remains constant at

the initial level R on all paddocks, irrespective of the chosen strategy, with the exception that (ii) he takes into account the extra income Δy (Equation 10.11) in a year after resting.

This means, a myopic farmer bases his decision on a very limited consideration of ecosystem dynamics: he only takes into account the short-term buffering function of reserve biomass, while neglecting all long-term ecological impact of the grazing management strategy chosen. Such a myopic farmer considers his income in year $t \geq 2$ to be

$$y(t) = r(t) \cdot \left\{ \begin{array}{ll} 1 & \text{if} \quad r(t) \leq \underline{r} \\ 1-\alpha & \text{if} \quad r(t) > \underline{r} \end{array} \right\} \cdot$$
$$\cdot \left[1 + w_R \cdot w_G \cdot \left(1 - \frac{R}{K}\right) \cdot r(t-1) \cdot \left\{ \begin{array}{ll} 0 & \text{if} \quad r(t-1) \leq \underline{r} \\ \alpha & \text{if} \quad r(t-1) > \underline{r} \end{array} \right\} \right] . \tag{10.14}$$

Since the myopic farmer neglects the long-term ecological impact of his grazing strategy, the functional form of how annual income $y(t)$ (Equation 10.14) depends on actual rainfall and on the chosen strategy, remains constant over time. Furthermore, since precipitation is independent and identically distributed in each year, and the strategy is constant, the mean $\mu_{y(t)}$ and standard deviation $\sigma_{y(t)}$ of the annual income $y(t)$ for $t \geq 2$ are also constant over time.

In order to be able to express the expected instantaneous utility in any year t in terms of the mean and the standard deviation of that year's income, we approximate the probability density function of annual income by a log-normal distribution with the same mean and standard deviation. Using the specification (10.13) of the Bernoulli utility function $u(y)$, expected instantaneous utility is given by the following explicit expression (see Appendix A10.2):

$$\mathcal{E}\, u(y(t)) = \frac{\mu_{y(t)}^{1-\rho} \left(1 + \sigma_{y(t)}^2 / \mu_{y(t)}^2\right)^{-\rho\,(1-\rho)/2} - 1}{1-\rho}. \tag{10.15}$$

The indifference curves of the farmer's expected instantaneous utility function can be drawn in the mean–standard deviation space. Figure 10.1 shows such a set of indifference curves for a given degree ρ of relative risk aversion. The indifference curves are increasing and convex if the standard deviation is sufficiently small compared to the mean, i.e. for $(\mu_y/\sigma_y)^2 > 1 + \rho$ (see Appendix A10.3). The slope of the indifference curves is increasing in the degree of relative risk aversion ρ (see Appendix A10.3). In particular, the indifference curves are horizontal lines for risk-neutral farmers, i.e. for $\rho = 0$.

Formally, the decision problem to be solved by a myopic farmer is to choose a grazing management strategy $(\alpha, \underline{r})$ such as to maximize U (Equation 10.12) subject to Conditions (10.7), (10.14), (10.15). In the context of semi-arid

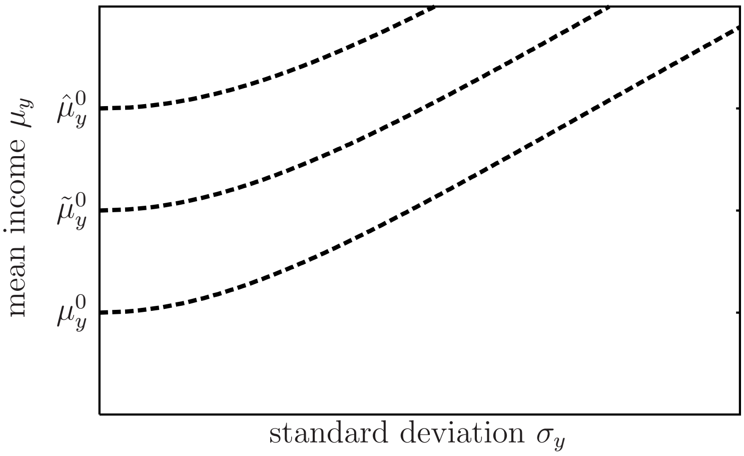

Figure 10.1 A set of indifference curves of the risk averse farmer in the mean-standard deviation space for log-normally distributed incomes and constant relative risk aversion $\rho = 1$.

rangelands, the growth rate of reserve biomass is small, i.e. $w_R \ll 1$. In Appendix A10.4 we show that under this condition the farmer's decision problem effectively becomes

$$\max_{(\alpha,\underline{r})} \; \mu_y(\alpha,\underline{r}) \cdot \left(1 + \sigma_y^2(\alpha,\underline{r})/\mu_y^2(\alpha,\underline{r})\right)^{-\rho/2} \;, \tag{10.16}$$

where the effective mean and standard deviation of income are

$$\mu_y(\alpha,\underline{r}) = \left[\mu_r - \alpha \int_{\underline{r}}^{\infty} r\, f(r)\, dr\right] \cdot \left[1 + \alpha\, \omega \int_{\underline{r}}^{\infty} r\, f(r)\, dr\right] \tag{10.17}$$

$$\sigma_y(\alpha,\underline{r}) = \sqrt{\sigma_r^2 + 2\,\alpha\,\mu_r \int_{\underline{r}}^{\infty} r\, f(r)\, dr - \alpha^2 \left[\int_{\underline{r}}^{\infty} r\, f(r)\, dr\right]^2 - \alpha\,(2-\alpha) \int_{\underline{r}}^{\infty} r^2\, f(r)\, dr} \cdot \sqrt{1 + 2\,\alpha\,\omega \int_{\underline{r}}^{\infty} r\, f(r)\, dr} \;, \tag{10.18}$$

with $\omega = w_R \cdot w_G \cdot (1 - R/K)/(1+\delta)$. We analyze this decision problem in the following.

10.4 Results

The analysis proceeds in three steps (Results 10.1, 10.2 and 10.3 below): First, we analyze the optimization problem of a risk averse myopic farmer who faces a trade-off between strategies which yield a high mean income at a high standard deviation, and strategies which yield a low mean income at a low standard deviation. Second, we analyze the long-term consequences of different grazing management strategies on the ecological-economic system. In particular, we study how the long-term development of the mean reserve biomass and the mean income depend on the strategy. Finally, we put the two parts of the analysis together and derive conclusions about how the long-term sustainability of the short-term optimal strategy depends on the farmer's degree of risk aversion.

10.4.1 Feasible Strategies and Income Possibility Set

To start with, we define the *income possibility set* as the set of all effective mean incomes and standard deviations of income $(\mu_y(\alpha, \underline{r}), \sigma_y(\alpha, \underline{r})) \in (0, \infty) \times [0, \infty)$, which are attainable by applying a feasible management rule $(\alpha, \underline{r}) \in [0, 1] \times [0, \infty)$. These are given by Equations (10.17) and (10.18). Figure 10.2 shows the income possibility set for particular parameter values.

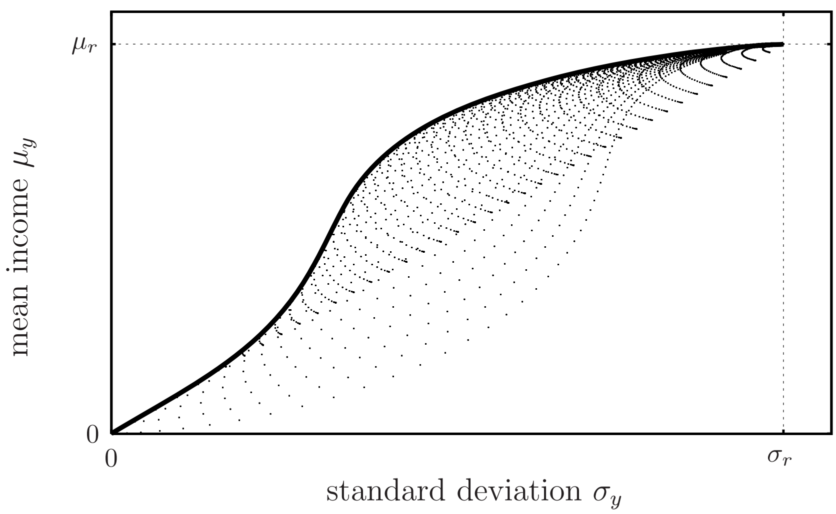

Figure 10.2 The set of all means μ_y and standard deviations σ_y of the farmer's income y, each point denoting a separate strategy, as well as the income possibility frontier (thick line). Parameter values are $\mu_r = 1.2$, $\sigma_r = 0.7$ and $\omega = 0.14$.

The figure provides one important observation: there exist inefficient strategies, i.e. feasible strategies that yield the same mean income, but with a higher standard deviation (or: the same standard deviation, but with a lower mean) than others. These strategies can be excluded from the set of strategies from which the optimum is chosen by a risk averse and non-satiated decision maker. In the following, we thus focus on the efficient strategies, which generate the *income possibility frontier* (Figure 10.2, thick line):

Definition 10.3
The *income possibility frontier* is the set of expected values μ_y and standard deviations σ_y of income for which the following conditions hold:

1. (μ_y, σ_y) is in the income possibility set, i.e. it is feasible.
2. There is no $(\mu'_y, \sigma'_y) \neq (\mu_y, \sigma_y)$ in the income possibility set with $\mu'_y \geq \mu_y$ and $\sigma'_y \leq \sigma_y$.

The question at this point is, 'What are the grazing management strategies $(\alpha, \underline{r})$ that generate the income possibility frontier?' We call these strategies *efficient*.

Lemma 10.1
The set of efficient strategies has the following properties.

- *Each point on the income possibility frontier is generated by exactly one (efficient) strategy.*
- *There exists* $\Omega \subseteq [0, \infty)$*, such that the set of efficient strategies is given by* $(\alpha^*(\underline{r}), \underline{r})$ *with*

$$\alpha^*(\underline{r}) = \frac{\int_{\underline{r}}^{\infty} r\,(r - \underline{r})\, f(r)\, dr}{\int_{\underline{r}}^{\infty} r\,(r - \underline{r}/2)\, f(r)\, dr} \quad \textit{for all} \quad \underline{r} \in \Omega. \tag{10.19}$$

- $\alpha^*(\underline{r})$ *has the following properties:*

$$\alpha^*(0) = 1, \quad \lim_{\underline{r} \to \infty} \alpha^*(\underline{r}) = 0, \quad \textit{and} \quad \frac{d\alpha^*(\underline{r})}{d\underline{r}} < 0 \quad \textit{for all} \quad \underline{r} \in \Omega.$$

Proof: see Appendix A10.5.

Figure 10.3 illustrates the lemma. Whereas the set of feasible strategies is the two-dimensional area bounded by $\underline{r} = 0$, $\alpha = 0$, $\alpha = 1$, the set of efficient strategies, as given by Equation (10.19), is a one-dimensional curve. Thus, the efficient strategies are described by only one parameter, $\underline{r}$, while the other parameter α is determined by $\alpha = \alpha^*(\underline{r})$ (Equation 10.19). Alternatively, the

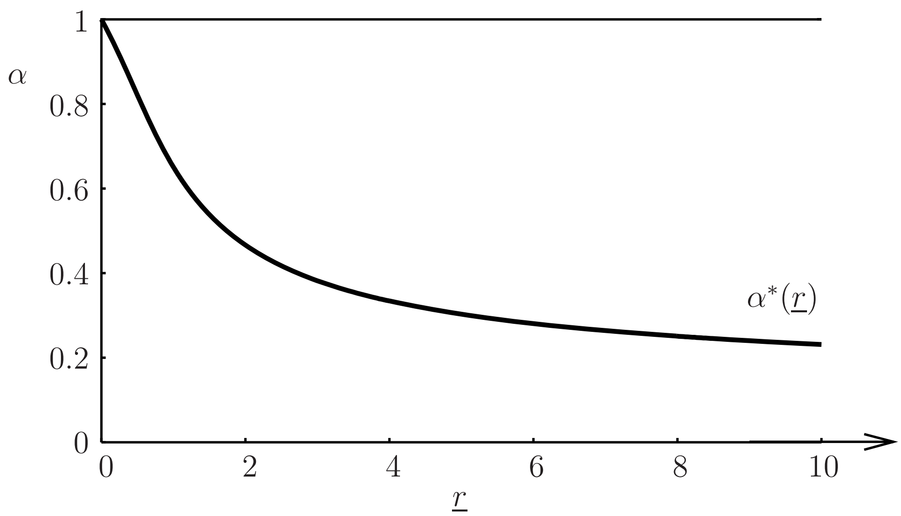

Figure 10.3 The set of feasible strategies is given by the whole area $\alpha \in [0,1]$, $\underline{r} \in [0,\infty)$. The set of efficient strategies for parameters $\mu_r = 1.2$ and $\sigma_r = 0.7$ is the curve.

inverse function of Equation (10.19) – which exists by Lemma 10.1 – specifies the efficient rain threshold $\underline{r}$ as a function of the fraction α of resting. The curve $\alpha^*(\underline{r})$ is downward sloping: With a higher rain threshold $\underline{r}$, i.e. if resting only takes place in years with higher precipitation, the efficient fraction $\alpha^*(\underline{r})$ of rested paddocks is smaller. In other words, for efficient strategies, a higher rain threshold $\underline{r}$ does not only mean that the condition for resting is less likely to be fulfilled, but also that a smaller fraction α^* of paddocks is rested if resting takes place. Hence, if an efficient strategy is characterized by a smaller $\underline{r}$, and, consequently, by a larger $\alpha^*(\underline{r})$, we call it *more conservative*.

Knowledge of the efficient strategies allows us to characterize the income possibility frontier, and to establish a relationship between efficient grazing management strategies and the resulting means and standard deviations of income.

Lemma 10.2
The farmer's expected income in the first grazing period, $\mu_y(\alpha, \underline{r})$ (Equation 10.17), is increasing in $\underline{r}$ for all efficient strategies:

$$\frac{d\,\mu_y(\alpha^*(\underline{r}), \underline{r})}{d\underline{r}} > 0 \quad \textit{for all } \underline{r} \in \Omega.$$

The extreme strategies, $\underline{r} = 0$ and $\underline{r} \to \infty$, lead to expected incomes of $\mu_y(\alpha^*(0), 0) = 0$ and $\lim\limits_{\underline{r}\to\infty} \mu_y(\alpha^*(\underline{r}), \underline{r}) = \mu_r$.

Proof: see Appendix A10.6.

For all efficient strategies a higher rain threshold $\underline{r}$ for resting, i.e. a less conservative strategy, implies a higher mean income. Whereas no resting, $\underline{r} \to \infty$ (opportunistic strategy), leads to the maximum possible mean income of μ_r, the opposite extreme strategy, $\underline{r} = 0$ (no grazing at all), leads to the minimum possible income of zero. Overall, a change in the grazing management strategy affects both, the mean income and the standard deviation of income.

Lemma 10.3
The income possibility frontier has the following properties:

- *The income possibility frontier has two corners:*
  - *The southwest corner is at $\sigma_y = 0$ and $\mu_y = 0$. At this point, the income possibility frontier is increasing with slope μ_r/σ_r.*
  - *The northeast corner is at $\sigma_y = \sigma_r$ and $\mu_y = \mu_r$. At this point, the income possibility frontier has a maximum and its slope is zero.*
- *In between the two corners, the income possibility frontier is increasing and located above the straight line from one corner to the other. It is S-shaped, i.e. from southwest to northeast there is first a convex segment and then a concave segment.*

Proof: see Appendix A10.7.

Figure 10.2 illustrates the lemma. With no resting at all (northeast corner of the income possibility frontier), the farmer obtains the highest possible mean income ($\mu_y = \mu_r$), but also faces the full environmental risk ($\sigma_y = \sigma_r$). Conversely, with the most conservative strategy, i.e. no grazing at all (southwest corner of the income possibility frontier), the farmer can completely eliminate his income risk ($\sigma_y = 0$), but also cannot expect any income ($\mu_y = 0$). The property, that the income possibility frontier is increasing, suggests that resting acts like an insurance for the farmer. This means, by choosing a more conservative grazing management strategy, the farmer can continuously decrease his risk (standard deviation) of income, but only at the price of a decreased mean income. Thus, there is an insurance value associated with choosing a more conservative strategy (Baumgärtner and Quaas 2005).

10.4.2 Optimal Myopic Strategy

The optimal myopic strategy is obtained by solving Problem (10.16), and results from both the farmer's preferences (Figure 10.1) and the income possibility frontier (Figure 10.2). In mean–standard deviation space, it is determined by the mean μ_y^* and the standard deviation σ_y^*, at which the indifference curve is tangential to the income possibility frontier (Figure 10.4). It turns out that the optimal strategy is uniquely determined.

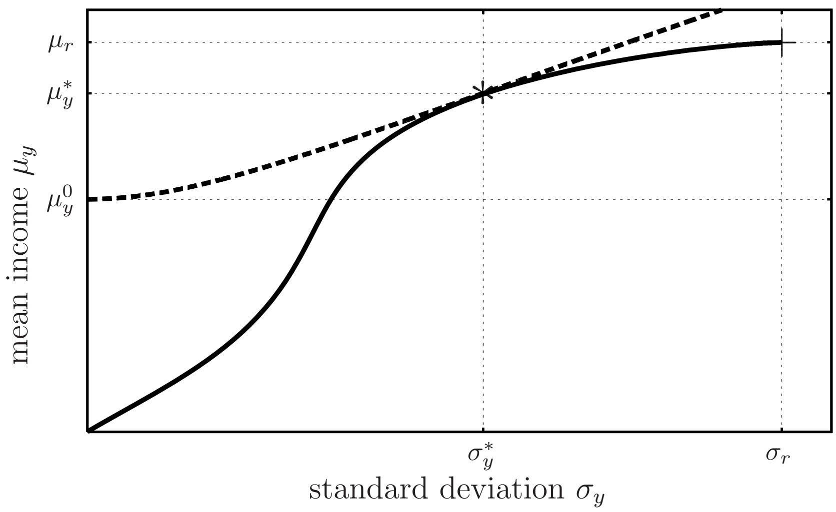

Figure 10.4 The optimum for a risk averse farmer ($\rho = 5.5$, denoted by $*$) and a risk-neutral farmer ($\rho = 0$, denoted by $+$).

Lemma 10.4

(i) If $(\mu_r/\sigma_r)^2 > 1 + \rho$, the optimum (μ_y^*, σ_y^*) is unique.[9]

(ii) For $\rho > 0$, the optimum is an interior solution with $0 < \mu_y^* < \mu_r$ and $0 < \sigma_y^* < \sigma_r$. For $\rho = 0$, the optimum is a corner solution with $\mu_y^* = \mu_r$ and $\sigma_y^* = \sigma_r$.

Proof: see Appendix A10.8.

The optimal myopic strategy crucially depends on the degree of risk aversion. In the particular case of a risk-neutral farmer ($\rho = 0$), the strategy that yields the maximum mean, irrespective of the standard deviation associated with it, is chosen. The optimal grazing management strategy of such a risk-neutral farmer is the strategy without resting, i.e. with $\underline{r} = \infty$ (and, therefore, $\alpha = 0$). That is, he employs an opportunistic strategy.

If the farmer is risk averse, he faces a trade-off between expected income and variability of the income, because strategies that yield a higher mean income also display a higher variability of income. This leads to the following result, which is illustrated in Figures 10.4 and 10.5.

Result 10.1

A unique interior solution $(\alpha^*(\underline{r}^*), \underline{r}^*)$ to the farmer's decision problem (10.16), if it exists (see Lemma 10.4), has the following properties:

---

[9]This is a sufficient condition which is quite restrictive. A unique optimum exists for a much larger range of parameter values.

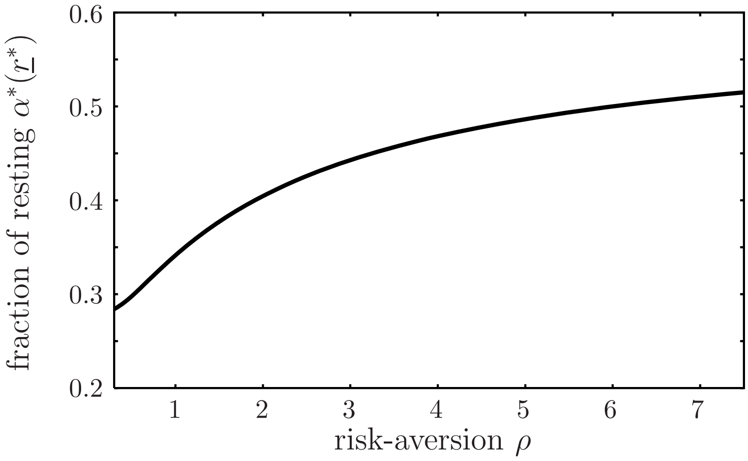

Figure 10.5 The rain threshold $\underline{r}^*$ of the optimal strategy as a function of the farmer's degree of risk aversion ρ. Parameter values are the same as in Figure 10.2.

(i) The more risk averse the farmer, the smaller are the mean μ_y^* and the standard deviation σ_y^* of his income.

(ii) The more risk averse the farmer, the more conservative is his grazing management strategy:

$$\frac{d\underline{r}^*}{d\rho} < 0 \qquad \textit{and} \qquad \frac{d\alpha^*}{d\rho} > 0. \tag{10.20}$$

Proof: see Appendix A10.9.

This means, a risk averse farmer chooses a grazing management strategy such as to obtain insurance from the ecosystem: by choosing a particular grazing management strategy the farmer will reduce his income risk, and carry the associated opportunity costs in terms of mean income foregone (the 'insurance premium'), to the extent that is optimal according to his degree of risk aversion.

10.4.3 Long-Term Impact of Grazing Management Strategies

To study the long-term ecological and economic impact of the grazing management strategy chosen on the basis of myopic optimization (Problem 10.16), we assume that the farmer continues to apply this strategy in every subsequent period. We compute the resulting probability distribution of income and reserve biomass over several decades in the future. This calculation covers all efficient strategies $(\alpha^*(\underline{r}), \underline{r})$. The results of the numerical computation[10] are shown

[10]Numerical details are given in Müller et al. (forthcoming).

in Figure 10.6, which enables the comparison of the long-term impacts, both in ecological and economic terms, of the different strategies that are efficient from the viewpoint of a myopic farmer. In this figure, the mean values $\mu_R(t)$

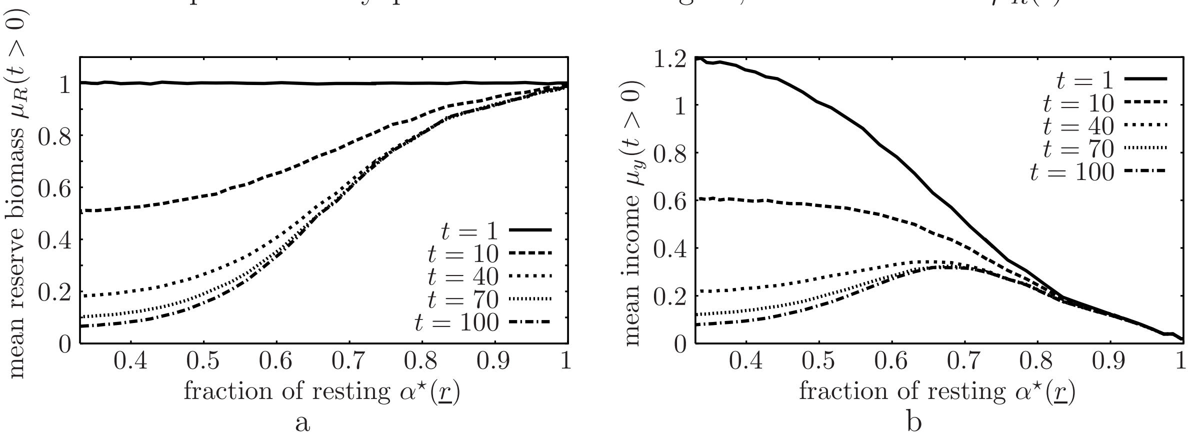

Figure 10.6 Relation between the grazing management strategy (given by the efficient fraction of resting $\alpha^\star(\underline{r})$) and (a) future mean reserve biomass $\mu_R(t >$ 0) (in units of initial reserve biomass), as well as (b) future mean income $\mu_y(t > 0)$ for different strategies on the income possibility frontier. Parameter values are $\mu_r = 1.2$, $\sigma_r = 0.7$, $I \cdot K = 8000$, $d = 0.15$, $w_G = 1.2$, $w_R = 0.2$, $c = 0.5$, $I \cdot R = 2400$.

of reserve biomass and $\mu_y(t)$ of income at different times t are plotted against the efficient fraction $\alpha^\star(\underline{r})$ of resting for different rain thresholds $\underline{r} \in \Omega$. The higher $\alpha^\star(\underline{r})$ is, the more conservative is the respective strategy. Interpreting Figure 10.6 leads to the following result (see Appendix A10.10 for a sensitivity analysis).

Result 10.2

For parameter values which characterize typical semi-arid rangelands (i.e. w_G, w_R, μ_r are small and c, σ_r are large) the long-term ecological and economic impact of a strategy $(\alpha^\star(\underline{r}), \underline{r})$ is as follows:

(i) The more conservative the strategy, the higher the mean reserve biomass $\mu_R(t)$ in the future:

$$\frac{d\,\mu_R(t)}{d\underline{r}} < 0 \quad \textit{and} \quad \frac{d\,\mu_R(t)}{d\alpha} > 0 \quad \textit{for all} \quad t > 1.$$

(ii) For high rain thresholds $\underline{r} > \hat{\underline{r}}$, the following holds: The more conservative the strategy, the higher the mean income $\mu_y(t)$ in the long-term future for $t > \hat{t}$:

$$\frac{d\,\mu_y(t)}{d\underline{r}} < 0 \quad \textit{and} \quad \frac{d\,\mu_y(t)}{d\alpha} > 0 \quad \textit{for all} \quad t > \hat{t} \quad \textit{and} \quad \underline{r} > \hat{\underline{r}},\ \alpha < \alpha^\star(\hat{\underline{r}}).$$

Result 10.2 states that the slope of the curves in Figure 10.6 is positive throughout, as far as reserve biomass is concerned; and is positive for small $\alpha^\star(\underline{r})$, i.e. for $\alpha^\star(\underline{r}) < \alpha^\star(\hat{\underline{r}})$, and $t > \hat{t}$, as far as income is concerned. The higher the fraction $\alpha^\star(\underline{r})$ of paddocks rested, i.e. the more conservative the strategy, the higher is the mean reserve biomass, if the same strategy is applied over the whole period. This effect is in line with intuition: the more conservative the strategy, the better is the state of the rangeland in the future. As far as income is concerned, the argument is less straightforward. In particular, the mean income in the first period is increasing in $\underline{r}$, i.e. decreasing in $\alpha^\star(\underline{r})$ (Lemma 10.2). A less conservative strategy yields a higher mean income in this period, since more livestock is kept on the rangeland. This holds for several periods in the near future (cf. the line for $t = 10$ in Figure 10.6b). However, in the long run (for $t > \hat{t} \approx 40$), the strong grazing pressure on the pasture leads to reduced reserve biomass growth and less forage production in the long-term future, compared to a more conservative strategy. As a result, mean income is smaller. This can be seen in Figure 10.6b: the curves are upward-sloping for sufficiently high $t \geq \hat{t}$ and sufficiently small $\alpha^\star(\underline{r})$. As can be seen in the figure, this effect becomes stronger in the long-term future: the curves are steeper for higher t.

Result 10.2 holds if the growth rates of the green and reserve biomass are low, the impact of grazing on the growth of the reserve biomass is high, and rainfall is low and highly variable. This is just the range of parameter values which is adequate for semi-arid rangelands, because these are fragile ecosystems which are highly susceptible to degradation if grazing pressure is high. For very robust ecosystems or very low stochasticity of rainfall, however, the result is not valid.

For a large fraction of resting, i.e. $\alpha^\star(\underline{r}) > \alpha^\star(\hat{\underline{r}})$, a more conservative strategy (i.e. a larger $\alpha^\star(\underline{r})$) leads to a lower mean income, not only in the first period (Lemma 10.2), but also in the future. In this domain of strategies, resting is already so high that the future gains in reserve biomass from additional resting do not outweigh the losses from lower stocking.

While Result 10.2 describes the dynamic long-term impact of different grazing management strategies, the following lemma analytically extends this result by specifying the steady-state mean values of reserve biomass and income. The steady-state mean value of reserve biomass is determined as the fixed point of the mean vegetation dynamics (according to Equations 10.1 and 10.2). The steady-state mean value of reserve biomass, in turn, determines the steady-state mean value of income.[11]

[11] These steady-state mean values represent the trend of the stochastic dynamics, but not the purely random part of the dynamics. The latter could lead, by chance, to irreversible extinction of the reserve biomass in the long-run even when a very conservative strategy is applied.

Lemma 10.5

1. *For an efficient strategy* $(\alpha^\star(\underline{r}), \underline{r})$ *the steady-state mean value of reserve biomass is*

$$\mu_R^{stst} = \max\left\{ K\, \frac{w_G\, w_R\, (\mu_R - c\, \mu_{y(1)}(\alpha^\star(\underline{r}), \underline{r})) - d}{w_G\, w_R\, (\mu_R - c\, \mu_{y(1)}(\alpha^\star(\underline{r}), \underline{r})) + d}, 0 \right\} \; , \tag{10.21}$$

and the steady-state mean value of income is

$$\mu_y^{stst} = \frac{\mu_R^{stst}}{R}\, \mu_{y(1)}(\alpha^\star(\underline{r}), \underline{r}) \; , \tag{10.22}$$

where $\mu_{y(1)}(\alpha^\star(\underline{r}), \underline{r})$ *is given by Equation* (10.8)*, and* R *is the initial value of reserve biomass.*

2. μ_R^{stst} *is monotonically decreasing in* $\underline{r}$*,*

$$\frac{d\mu_R^{stst}}{d\underline{r}} < 0 \; , \tag{10.23}$$

while μ_y^{stst} *assumes a maximum value at* $\hat{\underline{r}} > 0$*, such that*

$$\frac{d\mu_y^{stst}}{d\underline{r}} < 0 \quad \textit{for} \quad \underline{r} > \hat{\underline{r}} \; . \tag{10.24}$$

Proof: see Appendix A10.11.

For $\underline{r} > \hat{\underline{r}}$, we thus have established the following result: The more conservative the strategy, i.e. the lower $\underline{r}$ and the higher $\alpha^\star(\underline{r})$, the higher the steady-state mean reserve biomass and income in the long run.

As the final step in our analysis, we now relate this insight to the issue of sustainability of grazing management strategies. For the sake of this analysis, we understand *sustainability* in the following way.

Definition 10.4

A grazing management strategy $(\alpha, \underline{r})$ is called *sustainable*, if and only if it leads to strictly positive steady-state mean values of both reserve biomass and income, $\mu_R^{\text{stst}} > 0$ and $\mu_y^{\text{stst}} > 0$.

The notion of *sustainability*, while expressing an idea which seems obvious and clear at first glance, is notoriously difficult to define in an operational way. As a result, there are a multitude of different definitions of 'sustainability', which reveal different aspects and, at bottom, fundamentally different understandings of the term (see e.g. Klauer 1999, Neumayer 2003 and Pezzey 1992 for a detailed discussion). In the framework of our model, Definition 10.4 captures essential aspects of what has been called 'strong sustainability' (Pearce et al.

1990, Neumayer 2003). It comprises an ecological as well as an economic dimension, with mean reserve biomass as an ecological indicator and mean income as an economic indicator. It expresses the aspect of long-term conservation of an ecological-economic system in the sense that the steady-state mean values of both reserve biomass and income are strictly positive.[12] In contrast, an unsustainable strategy is one that leads to the collapse of the ecological-economic system, in the sense that the steady-state mean value of either reserve biomass or income (or both) is zero. Definition 10.4 constitutes a rather weak criterion of strong sustainability, by setting the minimum requirements with respect to the steady-state mean values of both reserve biomass and income at zero.[13] Yet, it enables a clear and unambiguous distinction between sustainable and unsustainable strategies in the following manner.

Lemma 10.6
If $c > 1 - d/(w_G\, w_R\, \mu_r)$, a strategy $(\alpha^\star(\underline{r}'), \underline{r}')$ exists, such that all efficient strategies which are less conservative (i.e. $\underline{r} > \underline{r}'$ and $\alpha^\star(\underline{r}) < \alpha^\star(\underline{r}')$) are unsustainable and all efficient strategies with $\underline{r} > 0$ that are more conservative (i.e. $\underline{r} < \underline{r}'$ and $\alpha^\star(\underline{r}) > \alpha^\star(\underline{r}')$) are sustainable.

Proof: see Appendix A10.12

If the impact of grazing on reserve biomass growth is very small, i.e. if $c < 1 - d/(w_G\, w_R\, \mu_r)$, all strategies are sustainable. Long-term degradation of the pasture is only a problem at all when the impact of grazing on the vegetation is high. In this case, there is a clear and unambiguous threshold between strategies that are conservative enough to be sustainable and strategies which are not. From Result 10.1, we know that the more risk averse a farmer is, the more conservative is his optimal myopic strategy. Combining this result with Lemma 10.6, we can now make a statement about the relation between a risk averse farmer's myopic decision and its long-term implications in terms of sustainability.

Result 10.3
If the uncertainty of rainfall, σ_r, is large and the impact of grazing c is not too large, a sufficiently risk averse myopic farmer will choose a sustainable grazing management strategy.

Proof: see Appendix A10.13.

[12]Under uncertainty, positive steady-state mean values do not mean that a sustainable strategy will actually yield positive values of reserve biomass and income. For, by chance, a sequence of rain events may occur which drives the reserve biomass to extinction. See Footnote 11.

[13]As an alternative, one could set minimum requirements at strictly positive values, representing e.g. the levels of 'critical natural capital' and 'subsistence income'. We have chosen zero for the sake of analytical clarity.

Result 10.3 sheds new light on the question 'How can one explain that people do behave in a sustainable way?' For, Result 10.3 suggests the following potential explanation. That a farmer A manages an ecosystem in a sustainable manner, while another farmer B does not, may be explained simply by a higher risk aversion of farmer A. In particular, it is not necessary to assume that farmer A has any kind of stronger preferences for future income or sustainability than farmer B. This result holds if (i) uncertainty is large and (ii) the impact of grazing is not too large. If uncertainty were small, it would only play a minor role in the decision making of the farmer. Hence, even a large risk aversion would not induce a myopic farmer to choose a conservative strategy. If, on the other hand, the impact of grazing were very high, the optimal strategy of even a very risk averse myopic farmer would not be conservative enough to ensure sustainability.

For a large standard deviation of rainfall and not too large grazing impacts, the model predicts a critical degree ρ' of risk aversion which separates the myopic farmers choosing a sustainable strategy from those choosing an unsustainable one. This critical degree of risk-aversion characterizes precisely that myopic farmer who chooses the strategy $(\alpha^\star(\underline{r}'), \underline{r}')$, which separates sustainable from unsustainable strategies (Result 10.1(ii) and Lemma 10.6). For the parameter values used in our numerical simulations (see the caption of Figure 10.6), this critical degree of risk aversion is $\rho' = 1.85$, which is well within the range of degrees of risk aversion commonly considered as reasonable (i.e. $\rho \leq 4$; see e.g. Gollier 2001).[14]

10.5 Conclusions and Discussion

We have developed an integrated dynamic and stochastic ecological-economic model of grazing management in semi-arid rangelands. Within this, we have analyzed the choice of grazing management strategies of a risk averse farmer, and the long-term ecological and economic impact of different strategies. We have shown that a myopic farmer who is sufficiently risk averse will choose a sustainable strategy, although he does not take into account long-term ecological and economic benefits of conservative strategies. The intuition behind this result is that a conservative strategy provides natural insurance for a risk averse farmer. In years with good rainfall the farmer does not fully exploit the carrying capacity of the farm. Due to the buffering function of the reserve biomass of vegetation he thereby can shift income to the next year with possibly worse conditions. The more risk averse the farmer is, the higher is the benefit from

[14] If the standard deviation of rainfall is small, or the grazing impact is very large, the threshold value of risk aversion exceeds this range of reasonable degrees of risk aversion.

this insurance function and the more conservative is his optimal strategy. A sufficiently risk averse farmer chooses a strategy which is conservative enough to be sustainable.

However, one should not conclude from our analysis that risk aversion is sufficient to ensure a sustainable development in semi-arid areas. This issue requires a variety of further considerations. First, one could adopt a more demanding sustainability criterion than we have used (cf. Definition 10.4). Second, we have focused on environmental risk resulting from the uncertainty of rainfall. Other forms of risk, e.g. uncertainty concerning property-rights, or the stability of social and economic relations in general, might generate a tendency in the opposite direction, and promote a less conservative and less sustainable management of the ecosystem (e.g. Bohn and Deacon 2000). Hence, in the face of different uncertainties, the net effect is not clear and has to be analyzed in detail. Third, additional sources of income (say from tourism) or the availability of financial services (such as savings, credits, or commercial insurance), constitute possibilities for hedging income risk. For farmers, all these are substitutes for obtaining natural insurance by conservative ecosystem management and, thus, may induce farmers to choose less conservative and less sustainable grazing management strategies (Quaas and Baumgärtner 2006). This becomes relevant as farmers in semi-arid regions are more and more embedded in world trade and have better access to global commodity and financial markets.

Our analysis addressed the context of grazing management in semi-arid rangelands. This system is characterized by a strong interrelation between ecology and economic use, which drives the results. While this is a specific ecological-economic system, the underlying principles and mechanisms of ecosystem functioning and economic management are fairly general. Hence, we believe that there are similar types of ecosystems managed for the services they provide, e.g. fisheries or other agro-ecosystems, to which our results should essentially carry over.

Appendix

A10.1 Mean and Standard Deviation of the First Year's Income

The rainfall r is log-normally distributed, i.e. the probability density function is

$$f(r) = \frac{1}{r\sqrt{2\pi s_r^2}} \exp\left(-\frac{(\ln r - m_r)^2}{2s_r^2}\right). \tag{A10.1}$$

The two parameters m_r and s_r can be expressed in terms of the mean μ_r and standard deviation σ_r, $m_r = \ln \mu_r - \frac{1}{2}\ln(1 + \sigma_r^2/\mu_r^2)$ and $s_r^2 = \ln(1 + \sigma_r^2/\mu_r^2)$.

Using the probability density function (A10.1) of rainfall and Equation (10.7) for the farmer's first year income, the expected value and the variance of the first year's income are easily calculated. The expected value is

$$\mu_{y(1)}(\alpha,\underline{r}) \;=\; \int_0^\infty y(1)\,f(r)\,dr = \int_0^{\underline{r}} r\,f(r)\,dr + (1-\alpha)\int_{\underline{r}}^\infty r\,f(r)\,dr = \mu_r - \alpha\int_{\underline{r}}^\infty r\,f(r)\,dr.$$

The variance is

$$\begin{aligned}
\sigma^2_{y(1)}(\alpha,\underline{r}) &= \int_0^\infty \left(y(1)-\mu_{y(1)}\right)^2\,f(r)\,dr = -\mu_y^2 + \int_0^{\underline{r}} r^2\,f(r)\,dr + (1-\alpha)^2\int_{\underline{r}}^\infty r^2\,f(r)\,dr \\
&= \sigma_r^2 + 2\,\alpha\,\mu_r\int_{\underline{r}}^\infty r\,f(r)\,dr - \alpha^2\left[\int_{\underline{r}}^\infty r\,f(r)\,dr\right]^2 - \alpha\,(2-\alpha)\int_{\underline{r}}^\infty r^2\,f(r)\,dr.
\end{aligned}$$

A10.2 Expected Utility Function

With the specification (10.13) of the farmer's Bernoulli utility function $u(y)$, and the assumption that income is log-normally distributed we get (using the notation $m_y = \ln\mu_y - \frac{1}{2}\ln\left(1+\sigma_y^2/\mu_y^2\right)$ and $s_y^2 = \ln\left(1+\sigma_y^2/\mu_y^2\right)$):

$$\begin{aligned}
\mathcal{E}\,u(y) &= \int_0^\infty \frac{y^{1-\rho}-1}{1-\rho}\,\frac{1}{y\sqrt{2\pi s_y^2)}}\exp\left(-\frac{(\ln y - m_y)^2}{2s_y^2}\right)dy \\
&\overset{z=\ln y}{=} \frac{1}{1-\rho}\left[\frac{1}{\sqrt{2\pi s_y^2}}\int_{-\infty}^\infty \exp\left((1-\rho)\,z\right)\,\exp\left(-\frac{(z-m_y)^2}{2s_y^2}\right)dz - 1\right] \\
&= \frac{\exp\left((1-\rho)\,\left(m_y + \frac{1-\rho}{2}\,s_y^2\right)\right)-1}{1-\rho} = \frac{\mu_y^{1-\rho}\,\left(1+\sigma_y^2/\mu_y^2\right)^{-\rho\,(1-\rho)/2}-1}{1-\rho}.
\end{aligned}$$

A10.3 Properties of the Indifference Curves

Each indifference curve intersects the μ_y-axis at $\sigma_y = 0$. The point of intersection, μ_0, is the certainty equivalent of all lotteries on that indifference curve. Hence, the indifference curve is the set of all $(\mu_y,\sigma_y)\in\mathbb{R}_+\times\mathbb{R}_+$ for which

$$\mu_y\,\left(1+\sigma_y^2/\mu_y^2\right)^{-\rho/2} = \mu_0. \qquad \text{(A10.2)}$$

The slope of the indifference curve is obtained by differentiating Equation (A10.2) with respect to σ_y (considering μ_y as a function of σ_y) and rearranging:

$$\frac{d\mu_y}{d\sigma_y} = \frac{\rho\,\sigma_y\,\mu_y}{(1+\rho)\,\sigma_y^2+\mu_y^2} > 0. \qquad \text{(A10.3)}$$

The curvature is obtained by differentiating this equation with respect to σ_y, inserting $d\mu_y/d\sigma_y$ again and rearranging

$$\frac{d^2\mu_y}{d\sigma_y^2} = \frac{d}{d\sigma_y}\frac{d\mu_y}{d\sigma_y} = \frac{\rho\,\mu_y\,(\mu_y^2 - (1+\rho)\,\sigma_y^2)(\sigma_y^2 + \mu_y^2)}{\left((1+\rho)\,\sigma_y^2 + \mu_y^2\right)^3}, \tag{A10.4}$$

which is positive, if and only if $\mu_y^2 > (1+\rho)\,\sigma_y^2$. Furthermore, the slope of the indifference curves increases with rising risk aversion,

$$\frac{d}{d\rho}\frac{d\mu_y}{d\sigma_y} = \frac{\sigma_y\,\mu_y\,\left(\sigma_y^2 + \mu_y^2\right)}{\left((1+\rho)\,\sigma_y^2 + \mu_y^2\right)^2} > 0.$$

A10.4 Effective Decision Problem

Using a monotonic transformation of U and employing $\sum_{t=2}^{\infty}(1+\delta)^{1-t} = 1/\delta$, the decision problem becomes

$$\max_{(\alpha,\underline{r})}\ \mu_{y(1)}\ \left(1 + \sigma_{y(1)}^2/\mu_{y(1)}^2\right)^{-\rho/2} + \frac{1}{\delta}\,\mu_{y(t)}\ \left(1 + \sigma_{y(t)}^2/\mu_{y(t)}^2\right)^{-\rho/2}\ . \tag{A10.5}$$

Since rainfall is independent and identically distributed in each year, we find from Equation (10.14) that mean income is

$$\mu_{y(t)} = \mu_{y(1)}\ \left[1 + \alpha\,w_R\,w_G\ \left(1 - \frac{R}{K}\right)\int_{\underline{r}}^{\infty} r\,f(r)\,dr\right]\ . \tag{A10.6}$$

When calculating the variance, we neglect terms of second order in w_R, since this is a very small number (see also the specification of parameters in the caption of Figure 10.6). With this simplification the variance is

$$\sigma_{y(t)} = \sigma_{y(1)}\ \sqrt{1 + 2\,\alpha\,w_R\,w_G\ \left(1 - \frac{R}{K}\right)\int_{\underline{r}}^{\infty} r\,f(r)\,dr}\ . \tag{A10.7}$$

Plugging this into the decision problem (A10.5) and again dropping terms of second order in the growth rate of the reserve biomass, we find

$$\begin{aligned}
&\mu_{y(1)}\ \left(1 + \sigma_{y(1)}^2/\mu_{y(1)}^2\right)^{-\rho/2} + \frac{1}{\delta}\,\mu_{y(t)}\ \left(1 + \sigma_{y(t)}^2/\mu_{y(t)}^2\right)^{-\rho/2} \\
&= \mu_{y(1)}\ \left(1 + \sigma_{y(1)}^2/\mu_{y(1)}^2\right)^{-\rho/2}\ \left[1 + \frac{1}{\delta}\left[1 + \alpha\,w_R\,w_G\ \left(1 - \frac{R}{K}\right)\int_{\underline{r}}^{\infty} r\,f(r)\,dr\right]\right] \\
&= \frac{1+\delta}{\delta}\,\mu_{y(1)}\ \left(1 + \sigma_{y(1)}^2/\mu_{y(1)}^2\right)^{-\rho/2}\ \left[1 + \alpha\,\frac{w_R\,w_G}{1+\delta}\ \left(1 - \frac{R}{K}\right)\int_{\underline{r}}^{\infty} r\,f(r)\,dr\right]
\end{aligned} \tag{A10.8}$$

Using the abbreviation $\omega = w_R\, w_G\, (1 - R/K)/(1+\delta)$, a monotonic transformation of the objective function, i.e. multiplication by $\delta/(1+\delta)$, and, once more, the approximation of dropping second-order terms in ω, one obtains the proposed result.

A10.5 Proof of Lemma 10.1

To find the efficient strategies, we first determine the strategies which minimize the standard deviation of income given the mean income. Out of these strategies those are efficient which maximize the mean income for a given standard deviation. Each point on the income possibility frontier is generated by exactly one efficient strategy, since the solution of the corresponding minimization problem is unique.

Equivalent to minimizing the standard deviation, we minimize the variance for a given mean income,

$$\min_{\alpha,\underline{r}} \sigma_y^2 \quad \text{s.t.} \quad \mu_y \geq \bar{\mu}_y,\ \alpha \in [0,1],\ \underline{r} \in [0,\infty). \tag{A10.9}$$

For a more convenient notation, we use the abbreviations

$$R_1(\underline{r}) = \int_{\underline{r}}^{\infty} r\, f(r)\, dr \quad \text{and} \quad R_2(\underline{r}) = \int_{\underline{r}}^{\infty} r^2\, f(r)\, dr. \tag{A10.10}$$

The Lagrangian for the minimization problem (A10.9) is

$$\begin{aligned}
\mathcal{L} &= \sigma_y^2(\alpha,\underline{r}) + \lambda\, [\mu_y(\alpha,\underline{r}) - \bar{\mu}_y] \\
&= \left[\sigma_r^2 + 2\,\alpha\,\mu_r\, R_1(\underline{r}) - \alpha^2\, R_1^2(\underline{r}) - \alpha\,(2-\alpha)\, R_2(\underline{r})\right] \cdot [1 + 2\,\alpha\,\omega\, R_1(\underline{r})] \\
&\qquad + \lambda\, [[\mu_r - \alpha\, R_1(\underline{r})] \cdot [1 + \alpha\,\omega\, R_1(\underline{r})] - \bar{\mu}_y]\, .
\end{aligned}$$

The first order condition with respect to $\underline{r}$ is

$$\begin{aligned}
&\alpha\,\underline{r}\, f(\underline{r})\ [-2\,(\mu_r - \alpha\, R_1(\underline{r})) + (2-\alpha)\,\underline{r}] \cdot [1 + 2\,\alpha\,\omega\, R_1(\underline{r})] \\
&\quad - \left[\sigma_r^2 + 2\,\alpha\,\mu_r\, R_1(\underline{r}) - \alpha^2\, R_1^2(\underline{r}) - \alpha\,(2-\alpha)\, R_2(\underline{r})\right] \cdot 2\,\omega\,\alpha\,\underline{r}\, f(\underline{r}) \\
&\quad = -\lambda\,\alpha\,\underline{r}\, f(\underline{r}) \cdot [1 + \alpha\,\omega\, R_1(\underline{r})] + \lambda\ [\mu_r - \alpha\, R_1(\underline{r})]\ \omega\,\alpha\,\underline{r}\, f(\underline{r}).
\end{aligned} \tag{A10.11}$$

The first order condition with respect to α is

$$\begin{aligned}
&[2\, R_1(\underline{r})\, (\mu_r - \alpha\, R_1(\underline{r})) - 2\,(1-\alpha)\, R_2(\underline{r})] \cdot [1 + 2\,\alpha\,\omega\, R_2(\underline{r})] \\
&\quad + \left[\sigma_r^2 + 2\,\alpha\,\mu_r\, R_1(\underline{r}) - \alpha^2\, R_1^2(\underline{r}) - \alpha\,(2-\alpha)\, R_2(\underline{r})\right] \cdot 2\,\omega\, R_1(\underline{r}) \\
&\quad = \lambda\, R_1(\underline{r}) \cdot [1 + \alpha\,\omega\, R_2(\underline{r})] - \lambda\ [\mu_r - \alpha\, R_1(\underline{r})]\ \omega\, R_1(\underline{r}).
\end{aligned} \tag{A10.12}$$

Canceling the common terms $\alpha\,\underline{r}\,f(\underline{r})$ in Equation (A10.11), and plugging the result into (A10.12) leads, with some rearranging, to

$$R_1(\underline{r})\,(2-\alpha)\,\underline{r} = 2\,(1-\alpha)\,R_2(\underline{r}) \quad \Leftrightarrow \quad \alpha^*(\underline{r}) = \frac{R_2(\underline{r}) - \underline{r}\,R_1(\underline{r})}{R_2(\underline{r}) - \frac{1}{2}\,\underline{r}\,R_1(\underline{r})}.$$

Re-inserting (A10.10) leads to (10.19), which is the unique solution of the first order conditions. $\sigma_y(\alpha^*(\underline{r}),\underline{r})$ is the minimum, since $\sigma_y(\alpha,\underline{r})$ is maximum at the corners $\alpha = 1$ (with $\underline{\rho} > 0$), or $\underline{\rho} = 0$ (with $\alpha < 1$), as can verified easily.

Equation (10.19) determines the set of strategies, which generate the minimum standard deviation for any given mean income. This set may include different strategies which lead to the same standard deviation, but different mean incomes. In such a case, we drop the strategy associated with the lower mean income, which is determined by $\alpha^*(\underline{r},\underline{r})$, where $\underline{r}$ is chosen from the appropriate subset $\Omega \subseteq [0,\infty)$ of feasible rain thresholds.

Turning to the properties of $\alpha^*(\underline{r})$, for $\underline{r} = 0$ the numerator and denominator of (10.19) are equal, hence $\alpha^*(0) = 1$. For $\underline{r} \to \infty$, we have, using L'Hospital's rule repeatedly, $\lim\limits_{\underline{r}\to\infty} \alpha^*(\underline{r}) = 0$. Numerical computations for a wide range of parameters (μ_r, σ_r) resulted in qualitatively the same curves $\alpha^*(\underline{r})$ as shown in Figure 10.3.

A10.6 Proof of Lemma 10.2

Plugging Equations (10.19) and (A10.10) into (10.17) and differentiating with respect to $\underline{r}$ yields:

$$\begin{aligned}\frac{d\,\mu_y(\alpha^*(\underline{r}),\underline{r})}{d\underline{r}} &= \left[-\frac{d\alpha^*(\underline{r})}{\underline{r}}\,R_1(\underline{r}) + \alpha^*(\underline{r})\,\underline{r}\,f(\underline{r})\right]\cdot[1 + 2\,\alpha\,\omega\,R_1 - \omega\,\mu_r] \qquad \text{(A10.13)}\\ &= \left[\frac{2\,R_1^2(\underline{r})\,R_2(\underline{r})}{(2\,R_2(\underline{r}) - \underline{r}\,R_1(\underline{r}))^2} + \alpha^{*2}(\underline{r})\,\underline{r}\,f(\underline{r})\right]\cdot[1 + 2\,\alpha\,\omega\,R_1 - \omega\,\mu_r] > 0,\end{aligned}$$

since, by assumption, $\omega\,\mu_r < 1$.

For $\underline{r} \to 0$, we have $\lim\limits_{\underline{r}\to 0} R_1(\underline{r}) = \mu_r$, $\lim\limits_{\underline{r}\to 0} R_2(\underline{r}) = \sigma_2^2 + \mu_r^2$, and $\alpha^*(0) = 1$. Inserting into equations (10.17) and (10.18) yields $\lim\limits_{\underline{r}\to 0} \mu_y(\alpha^*(\underline{r}),\underline{r}) = 0$ and $\lim\limits_{\underline{r}\to 0} \sigma_y(\alpha^*(\underline{r}),\underline{r}) = 0$.

For $\underline{r} \to \infty$, we have $\lim\limits_{\underline{r}\to\infty} R_1(\underline{r}) = 0$ and $\lim\limits_{\underline{r}\to\infty} R_2(\underline{r}) = 0$, and $\lim\limits_{\underline{r}\to\infty} \alpha^*(\underline{r}) = 0$. Inserting into equations (10.17) and (10.18) yields $\lim\limits_{\underline{r}\to\infty} \mu_y(\alpha^*(\underline{r}),\underline{r}) = \mu_r$ and $\lim\limits_{\underline{r}\to\infty} \sigma_y(\alpha^*(\underline{r}),\underline{r}) = \sigma_r$.

A10.7 Proof of Lemma 10.3

As shown in Appendix A10.6, $\lim_{\underline{r}\to\infty} \mu_y(\alpha^*(\underline{r}),\underline{r}) = \mu_r$ and $\lim_{\underline{r}\to\infty} \sigma_y(\alpha^*(\underline{r}),\underline{r}) = \sigma_r$. This is the northeast corner of the income possibility frontier, since $\mu_y = \mu_r$ is the maximum possible mean income (cf. Lemma 10.2). The slope of the income possibility frontier is

$$\frac{d\mu_y^{\text{ipf}}}{d\sigma_y} = \frac{d\mu_y(\alpha^*(\underline{r}),\underline{r})/d\underline{r}}{d\sigma_y(\alpha^*(\underline{r}),\underline{r})/d\underline{r}} = 2\,\sigma_y(\alpha^*(\underline{r}),\underline{r})\,\frac{d\mu_y(\alpha^*(\underline{r}),\underline{r})/d\underline{r}}{d\sigma_y^2(\alpha^*(\underline{r}),\underline{r})/d\underline{r}}.$$

From Appendix A10.5 we derive

$$\frac{d\sigma_y^2}{d\underline{r}} = -\lambda\,\frac{d\mu_y}{d\underline{r}} \tag{A10.14}$$

where λ is the costate-variable of the optimization problem (A10.9), which is determined by Equations (A10.11) and (A10.12),

$$-\lambda = \frac{-2\mu_y(\alpha^*(\underline{r}),\underline{r}) + (2-\alpha^\star(\underline{r}))\,\underline{r}\,(1+2\,\alpha^\star(\underline{r})\,\omega\,R_1(\underline{r})) - \frac{\sigma_y^2(\alpha^*(\underline{r}),\underline{r})\,2\,\omega}{1+2\,\alpha^\star(\underline{r})\,\omega\,R_1(\underline{r})}}{1+2\,\alpha^\star(\underline{r})\,\omega\,R_1(\underline{r}) - \omega\,\mu_r}\,.$$

Thus, we have

$$\frac{d\mu_y^{\text{ipf}}}{d\sigma_y} = \frac{2\,\sigma_y(\alpha^*(\underline{r}),\underline{r})}{-\lambda}. \tag{A10.15}$$

In particular for $\underline{r}\to\infty$, it is $\lim_{\underline{r}\to\infty}(-\lambda) = \frac{-2\,\mu_r + 2\,\underline{r} - \sigma_r^2\,2\,\omega}{1-\omega\,\mu_r} = \infty$. Hence,

$$\lim_{\underline{r}\to\infty} \frac{d\mu_y^{\text{ipf}}}{d\sigma_y} = 0.$$

For $\underline{r}\to 0$ both the mean income $\mu_y(\alpha^*(\underline{r},\underline{r}))$ and the standard deviation of income $\sigma_y(\alpha^*(\underline{r},\underline{r})$ vanish (cf. Appendix A10.6). Since both cannot be negative, this is the southwest corner of the income possibility frontier. At this point, the slope of the income possibility frontier is

$$\lim_{\underline{r}\to 0} \frac{d\mu_y^{\text{ipf}}}{d\sigma_y} = \lim_{\underline{r}\to 0} \frac{\mu_y(\alpha^*(\underline{r},\underline{r}))}{\sigma_y(\alpha^*(\underline{r}),\underline{r})} = \lim_{\underline{r}\to 0} \sqrt{\frac{(1-\alpha^*(\underline{r}))^2\,\mu_r^2\,(1+\alpha^*(\underline{r})\,\omega\,\mu_r)^2}{\sigma_r^2\,(1-\alpha^*(\underline{r}))^2\,(1+2\,\alpha^*(\underline{r})\,\omega\,\mu_r)}} = \frac{\mu_r}{\sigma_r},$$

neglecting terms of second order in ω. For $\underline{r} = 0$, and any given α, we have

$$\mu_y(\alpha,0) = [\mu_r - \alpha R_1(0)]\;[1+\alpha\,\omega\,R_1(0)] = (1-\alpha)\,\mu_r\,(1+\alpha\,\omega\,\mu_r)$$

$$\begin{aligned}\sigma_y^2(\alpha,0) &= \left[\sigma_r^2 + 2\,\alpha\,\mu_r\,R_1(0) - \alpha^2\,R_1^2(0) - \alpha\,(2-\alpha)\,R_2(0)\right]\;[1+2\,\alpha\,\omega\,R_1(0)]\\ &= (1-\alpha)^2\,\sigma_r^2\,(1+2\,\alpha\,\omega\,\mu_r),\end{aligned}$$

i.e., for small ω, the straight line between $(\mu_y, \sigma_y) = (0,0)$ $(\alpha = 1)$ and $(\mu_y, \sigma_y) = (\mu_r, \sigma_r)$ $(\alpha = 0)$ is always within the income possibility set. Since for $\underline{r} = 0$ the standard deviation is maximum for given mean income (cf. Appendix A10.5), the income possibility frontier is located above this straight line.

We have numerically determined the income possibility frontier for a large variety of parameters μ_r, σ_r, and ω. The results have provided strong evidence that under any set of parameters the income possibility frontier is divided into two domains: a convex domain for small σ_y and a concave domain for large σ_y. For very small σ_r, these two domains maybe separated by a jump in the income possibility frontier, such taht, in these extreme cases, the left borders of the respective income possibility sets inwardly curved to the right.

A10.8 Proof of Lemma 10.4

To prove part (i), we show that (a) the optimal indifference curve is convex over the whole range $\sigma_y \in [0, \sigma_r]$, and (b) the optimum is within the concave domain of the income possibility frontier.

Ad (a). Rearranging Equation (A10.2) yields the following expression for the optimal indifference curve (where μ_0^* is the certainty equivalent for the optimum)

$$\left(\frac{\sigma_y}{\mu_y}\right)^2 = \left(\frac{\mu_y}{\mu_0^*}\right)^{2/\rho} - 1. \tag{A10.16}$$

Inserting in the condition for the convexity of the indifference curve yields

$$\left(\frac{\mu_y}{\sigma_y}\right)^2 > 1 + \rho \quad \Leftrightarrow \quad \frac{\mu_y}{\mu_0^*} < \left(\frac{2+\rho}{1+\rho}\right)^{2/\rho}. \tag{A10.17}$$

By assumption, this condition is fulfilled for $\mu_y = \mu_r$ on the indifference curve which intersects (μ_r, σ_r), i.e. which is below the optimal one. Since $\mu_y \leq \mu_r$ for all efficient strategies, this condition is fulfilled for all μ_y on the optimal indifference curve.

Ad (b). The minimum slope of the income possibility frontier in the convex domain (i.e. at the southwest border) is μ_r/σ_r (Lemma 10.3). The slope of the indifference curve at the optimum (μ_y^*, σ_y^*), however, is smaller,

$$1+\rho < \left(\frac{\mu_r}{\sigma_r}\right)^2 < \frac{\mu_r}{\sigma_r}\frac{\mu_y^*}{\sigma_y^*} \quad \Rightarrow \quad \frac{\rho}{1+(1+\rho)\,\frac{\sigma_y^{*2}}{\mu_y^{*2}}} < \frac{\mu_r}{\sigma_r}\frac{\mu_y^*}{\sigma_y^*} \quad \Leftrightarrow \quad \frac{\rho\,\sigma_y^*\,\mu_y^*}{\mu_y^{*2} + (1+\rho)\,\sigma_y^{*2}} < \frac{\mu_r}{\sigma_r},$$

where the inequality $\mu_r/\sigma_r < \mu_y^*/\sigma_y^*$ holds as a consequence of Lemma 10.3, and the expression on the left hand side of the last inequality is the slope of the indifference curve at the optimum (cf. Equation A10.3). Hence, the optimum cannot be in the convex domain of the income possibility frontier.

Ad (ii). For $\rho = 0$, the indifference curves are horizontal lines. Hence, the maximum of the income possibility frontier, which is at the corner $(\mu_y, \sigma_y) = (\mu_r, \sigma_r)$, is the optimum.

For $\rho > 0$ corner solutions are excluded. At the corner $(\mu_y, \sigma_y) = (\mu_r, \sigma_r)$ the slope of the income possibility frontier is zero (Lemma 10.3), whereas the indifference curves have a positive slope, provided $\rho > 0$. At the corner $(\mu_y, \sigma_y) = (0, 0)$, the income possibility frontier is increasing with a slope μ_r/σ_r (Lemma 10.3), but the slope of the indifference curves is zero for $\sigma_r = 0$ (cf. Appendix A10.3).

A10.9 Proof of Result 10.1

We have shown that the unique optimum is in the concave domain of the income possibility frontier (Appendix A10.8), and that the slope of the farmer's indifference curves increases with ρ (Appendix A10.3). Thus, the optimal mean income μ_y^* decreases if ρ increases. Since for efficient strategies the mean μ_y^* is increasing in $\underline{r}$, the rain threshold $\underline{r}^*$ of the optimal strategy decreases if ρ increases.

A10.10 Sensitivity Analysis of Result 10.2

The aim of this Appendix is to show in a sensitivity analysis how the qualitative results shown in Figure 10.6 and stated in Result 10.2 depend on the parameters of the model. The sensitivity analysis was performed using a Monte Carlo approach, repeating the computations with multiple randomly selected parameter sets. We focused on three parameters, namely the growth parameter of green biomass w_G, the influence c of grazing on the growth of reserve biomass, and the standard deviation σ_r of rainfall. The other parameters either affect the outcomes in the same direction as the selected parameters (this is the case for the growth parameter of the reserve biomass w_R and the expected value of rainfall μ_r), or in the inverse direction (this is the case for the death rate of the reserve biomass d).[15] Hence their variation enables no further insights.

A sample size of $N = 20$ parameter sets was created according to the Latin Hypercube sampling method (Saltelli et al. 2000).[16] The three parameters were assumed to be independent uniformly distributed, with $0 \leq w_G \leq 5$, $0 \leq \sigma_r \leq 2.4$ and $0 \leq c \leq 1$, the upper bounds for w_G and σ_r are guesses which proved to be suitable. The respective simulation results were compared to the results shown in Figure 10.6. The following types of long-term dynamics of mean

[15] For the two parameters K and R, no substantial influence is to be expected: they just rescale the problem.

[16] This method, by stratifying the parameter space into N strata, ensures that each parameter has all proportions of its distribution represented in the sample parameter sets.

reserve biomass and mean income (distinct from those stated in Result 10.2) were found:[17]

(i) If the growth parameter of the green biomass w_G is very low, i.e. if $w_G \cdot w_R < d$, the reserve biomass is not able to persist at all. Keeping livestock is not possible, independent of the chosen grazing management strategy.

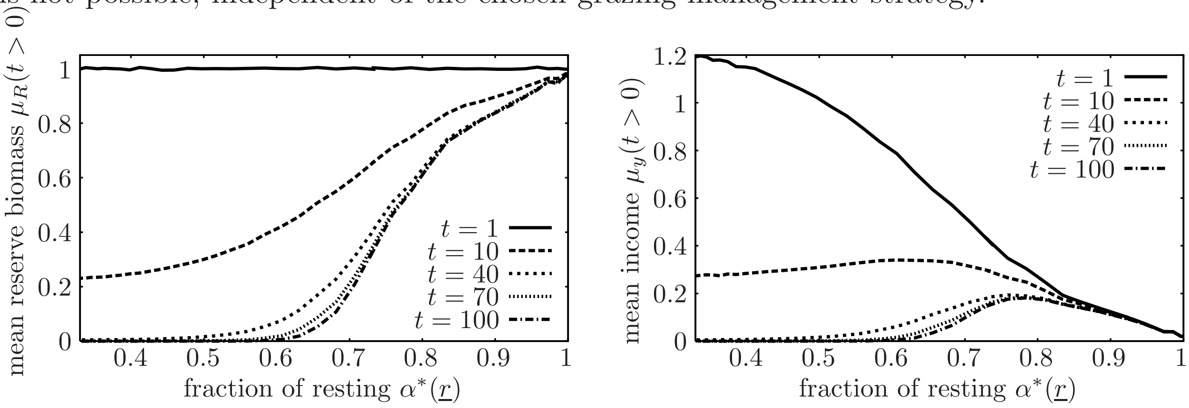

Figure A10.1 Parameter values are as in Figure 10.6, except for $c = 0.9$.

(ii) If the impact c of grazing on the growth of the reserve biomass is very high, the mean reserve biomass declines to zero in finite time, unless the grazing management strategy is very conservative. This is illustrated in Figure A10.1, where we have chosen $c = 0.9$.

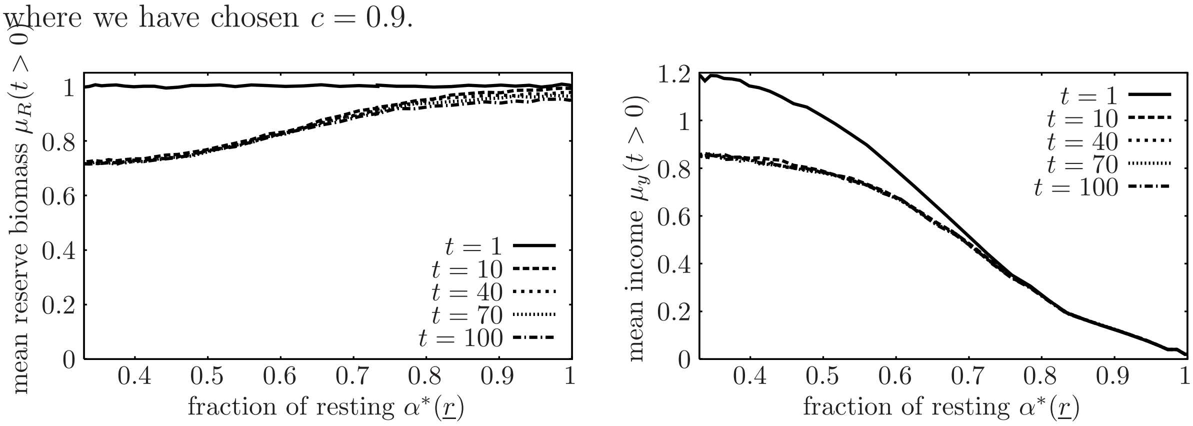

Figure A10.2 Parameter values are as in Figure 10.6, except for $w_G = 4$.

(iii) If the growth parameter of the green biomass is very high or the impact of grazing on the growth of the reserve biomass is very low, the future mean income is the higher the less conservative the strategy is, i.e. resting is not required to preserve the ecosystem. This is illustrated in Figure A10.2 for a very high growth rate of the biomass, $w_G = 4$. Qualitatively the same outcome arises for very low c (see also Müller et al. 2004).

[17] To illustrate them, additional calculations were done, where one parameter was chosen differently from the original parameter set of Figure 10.6 in each case.

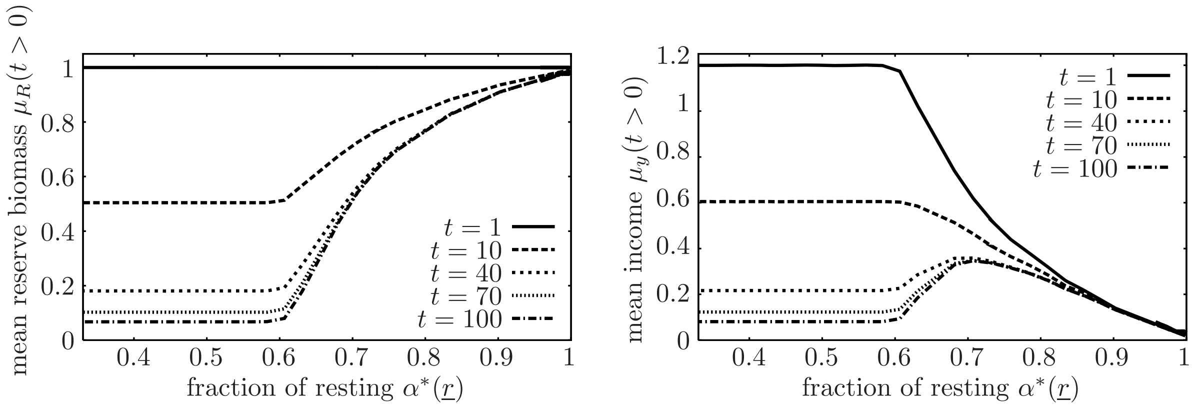

Figure A10.3 Parameter values are as in Figure 10.6, except for $\sigma_r = 0.05$.

(iv) If the standard deviation of rainfall σ_r is very small, resting is almost deterministic: for $\underline{r} > \mu_r$, resting will take place in hardly any year, such that mean reserve biomass μ_R and mean income μ_y are independent of the strategy. For $\underline{r} < \mu_r$, resting will take place in almost every year, i.e. the fraction $\alpha^*(\underline{r})$ of rested paddocks determines the outcome, as illustrated in Figure A10.3 for $\sigma_r = 0.05$.

A10.11 Proof of Lemma 10.5

In order to determine the steady-state mean value R^{stst} of the reserve biomass, we plug Equation (10.1) into Equation (10.2) an take the expected value on both sides of the resulting equation. In the long-term, the expectation value of R^i_t and R^i_{t+1} are the same and equal to R^{stst}. Given that in the long-term each camp will be rested with equal probability, we derive

$$d\, R^{\text{stst}} \left(1 + \frac{R^{\text{stst}}}{K}\right) = w_R\, w_G\, R^{\text{stst}} \left(1 - \frac{R^{\text{stst}}}{K}\right) \left(\mu_r - c\, \mu_{y(1)}(\alpha^\star(\underline{r}, \underline{r})\right).$$

This equation is solved by $R^{\text{stst}} = 0$ and by

$$R^{\text{stst}} = K\, \frac{w_G\, w_R\, (\mu_R - c\, \mu_{y(1)}(\alpha^\star(\underline{r}), \underline{r})) - d}{w_G\, w_R\, (\mu_R - c\, \mu_{y(1)}(\alpha^\star(\underline{r}), \underline{r})) + d}. \tag{A10.18}$$

If it is positive, the last expression is the solution; otherwise $R^{\text{stst}} = 0$ is the solution, since the reserve biomass cannot become negative. It is easily confirmed that R^{stst} is monotonically decreasing in $\mu_{y(1)}$. With a very similar argument as in Lemma 10.2, it is shown that $\mu_{y(1)}$ is monotonically increasing in $\underline{r}$. Hence, R^{stst} is monotonically decreasing in $\underline{r}$.

Income in each year is given by $y(t) = R(t)/(I\, R) \sum_{i=1}^{I} x^i\, r$. Given that each camp is equally likely to be rested in the long-term, the long-term expected value of income is

$$\mu_y^{\text{stst}} = \frac{\mu_R^{\text{stst}}}{R}\, \mu_{y(1)}(\alpha^\star(\underline{r}), \underline{r}). \tag{A10.19}$$

The unique interior extremum for which $R^{\text{stst}} > 0$ is given by

$$\hat{\mu}_{y(1)} = \frac{w_G\, w_R\, \mu_r + d - \sqrt{2\, d\, (w_G\, w_R\, \mu_r + d)}}{c\, w_G\, w_R} \,. \qquad \text{(A10.20)}$$

Is is a maximum, since for both corners $\mu_{y(1)} = 0$ and $\mu_{y(1)} = \mu_r$ we have $\mu_y^{\text{stst}} = 0$. Since $\mu_{y(1)}$ is monotonically increasing in $\underline{r}$, a unique $\hat{\underline{r}}$ exists, for which $\mu_{y(1)} = \hat{\mu}_{y(1)}$.

A10.12 Proof of Lemma 10.6

If $c > 1 - d/(w_G\, w_R\, \mu_r)$, $\mu_R^{\text{stst}} = 0$ for $\underline{r} \to \infty$, by Lemma 10.5. That is, a strategy without resting is unsustainable. If, however, $\underline{r} \to 0$, $\mu_{y(1)} = 0$ (by Lemma 10.2). Hence, as $d < w_G\, w_R\, \mu_r$, the strategy with complete resting is sustainable. By Lemma 10.2 $\mu_{y(1)}$ is monotonically increasing with $\underline{r}$, which concludes the proof.

A10.13 Proof of Result 10.3

By Lemma 10.6, all strategies are sustainable if $c \leq \overline{c} = \frac{w_G\, w_R\, \mu_r - d}{w_G\, w_R\, \mu_r}$. Hence, even the strategy chosen by risk-neutral farmers is sustainable. The interesting case is $c > \overline{c}$. In that case, the strategy chosen by a risk-neutral farmer is unsustainable. What remains to be shown is that for sufficiently large σ_r and sufficiently small c, a ρ' exists, such that all farmers with risk aversion $\rho > \rho'$ will choose a sustainable strategy. A necessary and sufficient condition for this statement is that

$$\lim_{\rho \to \infty} \mu_y(\alpha^\star(\underline{r}^\star(\rho)), \underline{r}^\star(\rho)) < \frac{w_G\, w_R\, \mu_r - d}{c\, w_G\, w_R} \,, \qquad \text{(A10.21)}$$

where $((\alpha^\star(\underline{r}^\star(\rho)), \underline{r}^\star(\rho))$ is the optimal strategy for a myopic farmer with risk aversion ρ. For, if Condition (A10.21) holds, the strategy chosen by an infinitely risk averse farmer is sustainable (cf. Lemma 10.6). Condition (A10.21) is fulfilled, if (i) the right hand side is large enough and (ii) the left hand side is small enough. The right hand side is large, if c and d are small. The right hand side is small, if σ_r is large compared to μ_r. This has been shown in Appendix A10.7: if σ_r is large, the income-possibility frontier is very flat in its concave domain. Hence, the optimal μ_y is only slightly smaller than μ_r, and Condition (A10.21) is violated, unless c is very small. In Figure A10.4, the threshold degree of risk aversion is plotted against σ_r (left hand side) and c (right hand side). For both, low σ_r and high c, this threshold value exceeds plausible values of ρ. But for high σ_r and comparatively low c, the threshold value ρ' lies well within the range of degrees of risk aversion which are commonly considered as reasonable ($\rho \leq 4$; see, e.g., Gollier 2001).

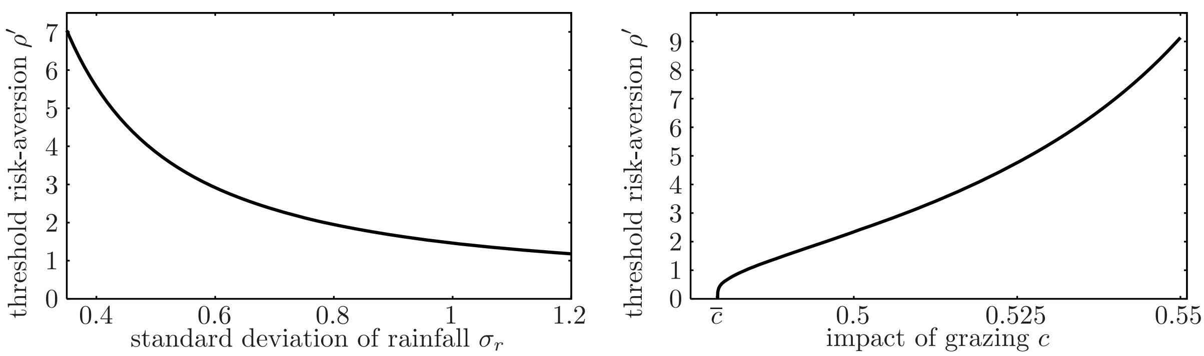

Figure A10.4 The threshold value of risk aversion, above which a myopic farmer chooses a sustainable strategy. On the left hand side plotted against the standard deviation σ_r of rainfall, on the right hand side plotted against the impact of grazing on vegetation. For $c \leq \overline{c} = \frac{w_G\, w_R\, \mu_r - d}{w_G\, w_R\, \mu_r}$, all strategies are sustainable (Lemma 10.6). The remaining parameter values are as in Figure 10.6.

11. Optimal Investment in Multi-Species Protection: Interacting Species and Ecosystem Health*

11.1 Introduction

The global loss of biodiversity currently proceeds at rates exceeding the natural rate of species extinction by a factor of 100 to 1000, mainly due to human disturbance of natural ecosystems (Watson et al. 1995b). As a response, in the past decades there have been an increasing number of policies targeted at the protection of endangered species, such as the U.S. Endangered Species Act (Brown and Shogren 1998). Only recently have these conservation policies come under scrutiny not only for their conservational effectiveness (Hoekstra et al. 2002, Shouse 2002) but also for their economic efficiency (Cullen et al. 2001, Dawson and Shogren 2001, Metrick and Weitzman 1996, 1998).

Under the U.S. Endangered Species Act the U.S. Fish and Wildlife Service, a division of the Department of Interior, lists species as endangered in the United States after (i) they have been suggested for listing by some individual or organization, public or private, (ii) scientific studies support the proposed listing, and (iii) no serious reasons against a listing emerge during a 60-day period for public comments. Listed species enjoy special protection from harm and must have official recovery plans created by the Fish and Wildlife Service. They are eligible for public spending on the federal and state levels. In 1995, there were 957 species listed as endangered in the United States and expenditures by federal and state agencies for all species recovery plans totalled US$ 280 million (Dawson and Shogren 2001).

As of the mid 1990s almost all endangered species had official recovery plans, but the expenditures were distributed rather unevenly among the different plans. Nearly 95% of the total reported spending by federal agencies were spent on about 200 vertebrate species, and only 5% were spent on about 800 invertebrate and plant species (Dawson and Shogren 2001). This has lead

*Previously published in *EcoHealth*, **1**(1), 101-110 (2004).

to the suggestion that the status of a species as 'charismatic megafauna' is a major factor in explaining the amount of funding for a recovery plan (Metrick and Weitzman 1996, 1998). While the number of species listed as endangered has almost doubled over the past decade – from 554 in 1989 up to 957 in 1995 – and total expenditure on species recovery programs has increased by a factor of almost seven – from US$ 44 million in 1989 to US$ 280 million in 1995 (Dawson and Shogren 2001) – only 13 species have actually recovered enough to warrant removal from the list (Shouse 2002).

One reason for the obviously poor performance of species recovery plans may be our poor understanding of the functioning of the natural ecosystems in which the target species live. The design of species recovery plans requires extensive knowledge of the species' life history and ecology (Bowles and Whelan 1994, MacMahon 1997). Yet, recent ecological surveys stress the large extent of uncertainty about the functioning of ecosystems (Brown et al. 2001, Holling et al. 1995, Loreau et al. 2001, Tilman 1997). Given this large uncertainty it is understandable that species recovery plans under the Endangered Species Act traditionally target single species, with the respective expenditure being highly species specific. Likewise, influential economic studies on optimal species protection plans for multi-species ecosystems assume that species are independent (Solow et al. 1993, Weitzman 1993, 1998).

However, considering species interactions is potentially important for the design of multi-species protection plans and to ensure the efficient allocation of limited conservation budgets (Wu and Bogess 1999). Here I show that taking species interaction into account makes a crucial difference for how to optimally allocate a given conservation budget. I conclude that effective species protection should go beyond targeting individual species, and consider species relations within whole ecosystems as well as overall ecosystem functioning. To make this conclusion operational I suggest to look at indicators of ecosystem health, which is a necessary prerequisite for successful species protection in situ.

11.2 Ecosystems and Species Extinction Risk

The formal framework used here follows and expands the one of Solow et al. (1993) and Weitzman (1998). Consider an ecosystem of $n \in \mathbb{N}$ different species. Each of them may be subject to stochastic extinction. Let p_i (with $i = 1, \ldots, n$) denote species i's survival probability, i.e. the probability that species i still exists after a time-period of T years.[1,2]

Let E_i (with $i = 1, \ldots, n$) be the status variable indicating whether species i will still be in existence after T years or whether it will have gone extinct:

$$E_i = \begin{cases} 1 & \text{if species } i \text{ survives,} \\ 0 & \text{if species } i \text{ becomes extinct.} \end{cases} \tag{11.1}$$

Due to the stochastic nature of extinction the variable E_i is a random variable. In general, the different E_i are not independent. The existence of certain species will influence the survival probabilities of others. This is most obvious for species that interact directly, for instance through a mutualistic relation (positive correlation between survival probabilities), competition for a common resource (negative correlation between survival probabilities), or a predator-prey relation. More generally, the relations among species in an ecosystem can be analyzed in terms of a trophic network. Such a food-web depicts the flow of food (measured in biomass) between the different species. The normalized flow between two species may be taken as a measure of the interaction strength between the two (Paine 1992). Food-web analysis permits to identify indirect interactions among species which are coupled through a food chain that comprises one or more intermediate nodes. Food-web analysis reveals the high degree of connectance and a complex pattern of species interactions even when looking at only a limited number of species in relatively few trophic groups (Elton 1927).

11.3 Human Appreciation of Species and Ecosystem Services

Individual species as well as entire ecosystems are valuable for humans for a number of reasons. Many species have direct use value as food, fuel, construction material, industrial resource or pharmaceutical substance (Farnsworth 1988, Plotkin 1988). More recently, it has been stressed that biodiversity, i.e. the set of all species, also has an important indirect use value in so far as entire ecosystems perform valuable services such as nutrient cycling, control of water

---

[1] The concepts of *extinction risk* of a population of species i, $1 - p_i$, and its *survival probability*, p_i, are equivalent measures of population viability (Burgman et al. 1993). Here, survival probabilities are used for the ease of interpretation. Another equivalent measure of population viability is its *expected lifetime*, τ_i. It is related to the survival probability p_i over a time-period T via the equation $p_i = \exp(-T/\tau_i)$ (Wissel et al. 1994). T is typically taken to be 10, 50 or 100 years in population viability analysis.

[2] On a more fundamental level the survival probabilities are determined by a number of factors, such as the species' population size, geographic range, age structure and spatial distribution (Lande 1993).

runoff, purification of air and water, soil regeneration, pollination of crops and natural vegetation, or partial climate stabilization (Daily 1997a, Mooney and Ehrlich 1997, Perrings et al. 1995b). These ecosystem services are essential to support the human existence on Earth. They can only be provided by more or less intact ecosystems and result from the complex – and, up to now, not well understood – interplay of many different species in these ecosystems (Holling et al. 1995, Tilman 1997).

Following Weitzman (1998) and Metrick and Weitzman (1998), the utility gained directly and indirectly from a multi-species ecosystem can be written as a sum of the direct utilities of all individual species, U_i ($i = 1, \ldots, n$), and the utility gained indirectly from the entire ecosystem through the ecosystem services provided collectively by all species, U_{ES}. In general, the utility of ecosystem services will be a function of the existence or non-existence of all species, $U_{ES} = U_{ES}(E_1, \ldots, E_n)$. Hence:

$$U = U_{ES}(E_1, \ldots, E_n) + \sum_{i=1}^{n} U_i. \tag{11.2}$$

For example, Weitzman (1998) specifies U_{ES} as the diversity of the set of all actually existing species. His diversity function provides an aggregate measure of the diversity of a set of species based on the pairwise dissimilarities among them (Weitzman 1992). This is in line with the idea that biodiversity may be taken as as a proxy for an ecosystem's capability of providing the valuable services described above (Holling et al. 1995, Loreau et al. 2001, Perrings et al. 1995b, Tilman 1997).

Because of the stochastic risk of species extinction a decision maker will consider not the utility, U, but the expected utility, $\mathcal{E}[U]$. With p_i as species i's survival probability the expected direct utility of that species is given by $p_i U_i$. Hence,

$$\mathcal{E}[U] = \mathcal{E}[U_{ES}(E_1, \ldots, E_n)] + \sum_{i=1}^{n} p_i U_i. \tag{11.3}$$

Specification of the function U_{ES} would require a detailed ecological model of how all the species in an ecosystems collectively provide certain ecosystem services. In order to keep matters simple I shall assume that the ecosystem provides all its services at full scale if, and only if, species 1 exists. The level of utility derived from ecosystem services then only depends on whether species 1 exists or not. Furthermore, in order to focus on species interaction in the ecosystem (instead of trade-offs on the utility side) I assume that all species have vanishing direct utility: $U_1 = \ldots = U_n = 0$. The value of all the different species, thus, is an indirect one and consists of their contribution to ecosystem functioning, and, in particular, of their support of species 1. Hence, the relevant objective function for making conservation decisions is

$$\mathcal{E}[U] = p_1 U_{ES}, \tag{11.4}$$

where U_{ES} is a positive constant. While species 1 thus plays a prime role, all the other species are potentially important, too, as their existence or absence may influence species 1's survival probability, p_1. I will address this latter point explicitly in Section 11.5 below.

11.4 Species Protection Plans and Optimal Allocation of a Conservation Budget

Consider now the economic decision problem of how to allocate a conservation budget among different species protection plans. The time structure of the problem is as follows. The decision about how to allocate the conservation budget is made today, and the corresponding species protection plans are enacted immediately. The result in terms of actual species survival or extinction is observed tomorrow (which means, more precisely, after the course of T years). The actual ecosystem situation tomorrow yields a certain utility, the expectation of which is the basis for today's decision.

For the moment, I will neglect species interaction, as it is done in the existing economic literature (Solow et al. 1993, Weitzman 1993, 1998). That is, in this section I will introduce the economic decision framework for the case that all n species are independent. I will then introduce species interaction in Section 11.5 below.

Following Weitzman (1998) assume that investment in some protection plan aimed at species i can enhance that species' survival probability p_i within certain limits:

$$\underline{p_i} \leq p_i \leq \overline{p_i} \qquad \text{with} \qquad \underline{p_i} \geq 0 \text{ and } \overline{p_i} \leq 1. \tag{11.5}$$

The probability $\underline{p_i}$ gives the 'down-risk' for species i's survival. This is the survival probability if no investment in protection is made. On the other hand, $\overline{p_i}$ indicates the 'up-risk' for species i. This is the maximum survival probability amenable for species i through the particular protection plan under consideration. In the extreme, $\underline{p_i} = 0$ and $\overline{p_i} = 1$. That is, without protection species i will become extinct for sure, but undertaking the protection plan at full scale will save it for sure. Any protection plan can also be undertaken at any level in between not-at-all and full-scale, leading to survival probabilities p_i which are on a continuum $\underline{p_i} \leq p_i \leq \overline{p_i}$.

Species protection plans are also costly. Suppose that out of an exogenously given and fixed budget $b > 0$ an amount $b_i \geq 0$ is spent on protecting species i. Then the following budget constraint holds:

$$\sum_{i=1}^{n} b_i \leq b \ . \tag{11.6}$$

Investment b_i in protecting species i will enhance the species' survival probability p_i according to a 'survival probability enhancement function', or 'enhancement function' for short:

$$p_i = P_i(b_i) \qquad \text{with } P_i(0) = \underline{p_i},\ P_i(b_i) \leq \overline{p_i} \text{ for all } b_i,\ P_i' \geq 0. \tag{11.7}$$

The qualifying properties state that without any investment species i's survival probability will stay at the lower bound, $\underline{p_i}$. On the other hand, species i's survival probability cannot exceed its upper bound, $\overline{p_i}$, no matter how much is invested in its protection. Generally, the more money is spent to enhance species i's survival probability the higher will p_i actually turn out to be. For example, Weitzman (1998) uses linear enhancement functions with

$$P_i(b_i) = \min\left\{\frac{b_i}{c_i}\left(\overline{p_i} - \underline{p_i}\right) + \underline{p_i}\ ,\ \overline{p_i}\right\}\ , \tag{11.8}$$

where the parameter $c_i > 0$ indicates the costs of enhancing the survival probability all the way from its lower bound $\underline{p_i}$ to its upper bound $\overline{p_i}$.

The economic decision problem can then be stated as follows: choose a budget allocation such as to maximize the expected utility function (11.3) subject to the budget constraint (11.6) and the feasible possibilities for survival probability enhancement as described by (11.7). Formally:

$$\begin{aligned}\text{maximize}_{\{b_i\}_{i=1,\ldots,n}}\ \mathcal{E}[U] \quad \text{s.t.} \quad & \sum_{i=1}^{n} b_i \leq b \quad \text{and} \\ & p_i = P_i(b_i) \text{ for all } i = 1, \ldots, n.\end{aligned} \tag{11.9}$$

This is a typical stochastic programming problem which is continuous in the b_i.

Weitzman (1998) has characterized the solution to problem (11.9) under the assumptions that (i) $\mathcal{E}[U_{ES}(E_1, \ldots, E_n)]$ is specified as the expected diversity of the set of all species, (ii) all species are independent and (iii) the enhancement functions are linear and given by (11.8). Obviously, with the simple objective function (11.4) the optimal solution, $\{b_i^*\}_{i=1,\ldots,n}$, is that the entire conservation budget is spent on species 1:[3]

$$b_1^* = b \quad \text{and} \quad b_i^* = 0 \text{ for } i = 2, \ldots, n. \tag{11.10}$$

11.5 Species Interaction

The formalization of problem (11.9) above, as well as the properties of its solution (11.10), rely heavily on the simplifying assumption of independent species. To illustrate how one could construct a more general framework for the case of interacting species let me introduce the effect of species interaction for a simple model ecosystem that consists of just two species ($n = 2$), and in which the existence of species 2 influences the survival probability of species 1 but not vice versa. For example, one could think of species 2 as a potential prey for species 1 (positive interaction), or as a predator of it (negative interaction). Focusing on $n = 2$ is not as restrictive as it may appear at first sight. For species 2 may be interpreted as 'all the rest of the ecosystem' besides species 1. In this interpretation, it then also appears plausible to assume that while species 2 influences species 1's survival probability, the reverse influence is negligible.

In this case the survival probability of species 1 ('target species') depends on the existence of species 2 ('support species'). Let $p_{1|E_2}$ denote the conditional survival probability of species 1 given the existence or non-existence of species 2. In particular, $p_{1|1}$ is the survival probability of species 1 if species 2 exists ($E_2 = 1$) and $p_{1|0}$ is the survival probability of species 1 if species 2 does not exist ($E_2 = 0$). The (unconditional) survival probability of species 1, taking into account that species 2 exists with probability p_2, is then given by

$$p_1 = p_{1|1}p_2 + p_{1|0}(1 - p_2). \tag{11.11}$$

An investment in a protection plan for species 1 will increase the conditional survival probability $p_{1|E_2}$:

$$p_{1|E_2} = P_{1|E_2}(b_1), \qquad \text{where } E_2 = \begin{cases} 1 & \text{if species 2 exists,} \\ 0 & \text{if species 2 is extinct.} \end{cases} \tag{11.12}$$

In particular, the existence of species 2 may be thought of as having an influence on the up and down risk for species 1, which also become conditional probabilities: $\underline{p_{1|E_2}}$ and $\overline{p_{1|E_2}}$. If the existence of species 2 has a positive influence on species 1 it seems natural to assume that

$$\underline{p_{1|1}} \geq \underline{p_{1|0}} \quad \text{and} \quad \overline{p_{1|1}} \geq \overline{p_{1|0}} \tag{11.13}$$

---

[3]For $c_1 < b$ the budget will not be completely exhausted by funding a full-scale protection plan for species 1. Since spending money on protecting other species would not increase utility under the objective function (11.3) the remaining budget, $b - b_1$, could either be left idle or allocated randomly among the other species.

with at least one inequality holding as a strict inequality. This is illustrated in Figure 11.1 which shows the feasible range of species 1's survival probability, $p_{1|E_2}$, conditional on the the absence ($E_2 = 0$) or existence ($E_2 = 1$) of species 2. The effect of a positive species interaction essentially is that it shifts the up-

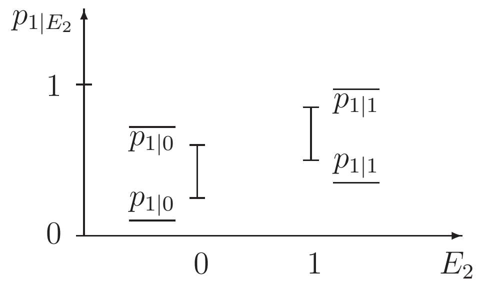

Figure 11.1 Feasible range of species 1's survival probability conditional on the the existence or absence of species 2, $p_{1|E_2}$ in the case of a positive influence.

risk and the down-risk for species 1, and therefore the entire feasible range of survival probabilities, upward.

Note that there is actually a number of ways, all consistent with condition (11.13), in which the existence of species 2 may have a positive influence on species 1's range of survival probabilities (Figure 11.2). One possibility (Fig-

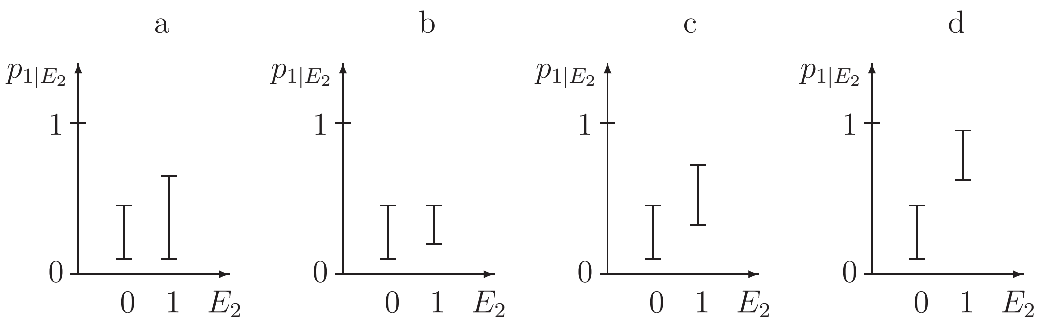

Figure 11.2 Different possibilities (a–d)of how the existence of species 2 may have a positive influence on species 1's feasible range of conditional survival probabilities.

ure 11.2a) is that under the positive influence of species 2 ($E_2 = 1$) the upper bound for species 1's conditional survival probability increases while the lower bound is not altered compared with a situation in which species 2 is absent ($E_2 = 0$). Or the lower bound for the conditional survival probabilities may increase while the upper bound is not altered (Figure 11.2b). Another possibility is that both the lower and upper bound increase such that the entire range of feasible conditional survival probabilities shifts upward (Figure 11.2c, d). This may happen in such a way that the range with and without existence of species 2 overlap (Figure 11.2c) or such that they do not overlap (Figure 11.2d). The latter case may be particularly relevant for evolutionary old ecosystems in which target species have coevolved with, and are well adapted in a special way to, their support species and ecosystem. For well adapted and specialized target species the existence of supporting species and ecosystems may have a larger effect on the species' survival probability than any protection plan aimed directly at that species.

If the existence of species 2 has a negative influence on species 1 one has:

$$\underline{p_{1|1}} \leq \underline{p_{1|0}} \quad \text{and} \quad \overline{p_{1|1}} \leq \overline{p_{1|0}}, \tag{11.14}$$

where at least one inequality holds as a strict inequality. Like in the case of positive interaction, condition (11.14) can be fulfilled in a variety of ways. And if the existence of species 2 does not have any influence on species 1 one has

$$\underline{p_{1|1}} = \underline{p_{1|0}} \quad \text{and} \quad \overline{p_{1|1}} = \overline{p_{1|0}}. \tag{11.15}$$

In this formal framework the economic decision problem of how to allocate a conservation budget among interacting species, now reads as follows:

$$\begin{aligned} \text{maximize}_{\{b_1,b_2\}}\ \mathcal{E}[U] \quad \text{s.t.} \quad & b_1 + b_2 \leq b, \\ & p_{1|E_2} = P_{1|E_2}(b_1) \text{ and } p_2 = P_2(b_2). \end{aligned} \tag{11.16}$$

11.6 Species Interaction and Optimal Allocation of the Conservation Budget

Species interaction can make a big difference for how to optimally allocate a conservation budget among different species protection plans. This is illustrated in this section by the example of a concrete parameterization of species interaction based on the formal framework developed in the previous section.

According to the objective function (11.4), if species 1 exists the utility is U_{ES}, and it is zero otherwise. Assume that the feasible range of survival probabilities for species 2 comprises the entire interval $[0, 1]$, i.e. $\underline{p_2} = 0$ and $\overline{p_2} = 1$. The feasible range of survival probabilities for species 1 is contingent upon the existence of species 2 and, furthermore, depends on the type and strength of influence of species 2 on species 1:

$$\begin{aligned} \text{with species 2 } (E_2 = 1): \quad & \frac{1}{5}(2 + 2\kappa) \ \leq \ p_{1|1} \ \leq \ \frac{1}{5}(3 + 2\kappa), \\ \text{without species 2 } (E_2 = 0): \quad & \frac{2}{5} \ \leq \ p_{1|0} \ \leq \ \frac{3}{5}, \end{aligned} \tag{11.17}$$

where $\kappa \in [-1, +1]$ parameterizes the influence of species 2 on species 1's survival probability conditional on the existence of species 2. With $\kappa = 0$ the two are independent and the existence of species 2 does not make any difference for the range of survival probabilities of species 1. With $\kappa > 0$ (< 0) species 2 has a positive (negative) influence on species 1's survival probability. The entire range of feasible survival probabilities is shifted upwards (downwards). Figure 11.3 shows the feasible range of survival probabilities for species 1 with and without existence of species 2 depending on the interaction strength κ.

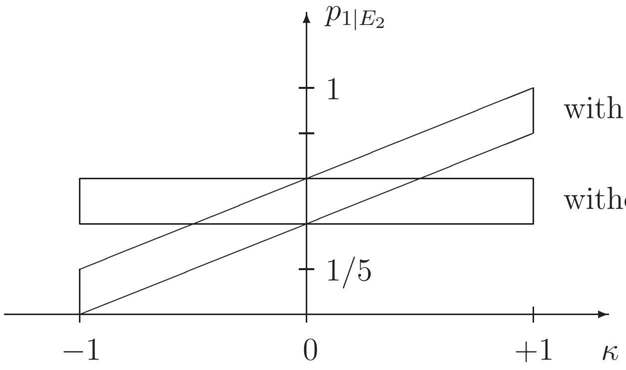

Figure 11.3 Feasible range of species 1's survival probability conditional on the existence or non-existence of species 2, $p_{1|E_2}$, depending on the interaction strength κ.

Assume that the total conservation budget is $b = 1$ and the enhancement functions for both species are as follows:

$$P_{1|E_2}(b_1) = \begin{cases} \frac{1}{5}\sqrt{b_1} + \frac{2}{5} & \text{without species 2 } (E_2 = 0) \\ \frac{1}{5}\sqrt{b_1} + \frac{2}{5}(1+\kappa) & \text{with species 2 } (E_2 = 1) \end{cases}, \tag{11.18}$$

$$P_2(b_2) = \sqrt{b_2}\ . \tag{11.19}$$

Note that the enhancement functions for both species exhibit strictly decreasing returns. With a budget of $b = 1$ and enhancement possibilities as specified here the economically viable survival probabilities for both species are within the feasible range described by (11.17). The budget of $b = 1$ allows either a full scale conservation project for species 1, or a full scale project for species 2, or projects for both of them at less than full scale. Spending the entire budget on species 1 allows to increase its survival probability, for given interaction strength κ and contingent on the existence or non-existence of species 2, from its lower bound to its upper bound. Similarly, spending the entire budget on species 2 allows to increase its survival probability from its lower bound to its upper bound. As these bounds for species 2 are given by 0 and 1, the size of the budget ($b = 1$) and the particular form of enhancement function (11.19) allows full control over species 2. With $b_2 = 0$, species 2 will be extinct for sure; with $b_2 = 1$, it will exist for sure; and for all levels $0 < b_2 < 1$, it will exist with probability $p_2 = \sqrt{b_2}$. This simple setting focuses on the influence of the interaction between the two species on how to split up the total budget between the two in order to maximize species 1's expected survival probability.

With $b = 1$ and b_2 as the expenditure on species 2, the remaining budget of $b_1 = 1 - b_2$ can be spent on species 1. The expected utility is given as the survival probability of species 1 times the utility derived from it. With (11.4) and (11.11):

$$\mathcal{E}[U] = p_1 U_{ES} = \left[p_{1|1}p_2 + p_{1|0}(1-p_2)\right] U_{ES}, \tag{11.20}$$

where $p_{1|1}$, $p_{1|0}$, and p_2 depend on the expenditures b_1 and b_2 according to the enhancement functions (11.18) and (11.19). With $b_1 = 1 - b_2$ one has

$$\begin{aligned}\mathcal{E}[U] &= \left[\left(\frac{1}{5}\sqrt{1-b_2} + \frac{2}{5}(1+\kappa)\right)\sqrt{b_2} + \left(\frac{1}{5}\sqrt{1-b_2} + \frac{2}{5}\right)(1-\sqrt{b_2})\right] U_{ES} \\ &= \left[\frac{2}{5} + \frac{2}{5}\kappa\sqrt{b_2} + \frac{1}{5}\sqrt{1-b_2}\right] U_{ES}. \end{aligned} \tag{11.21}$$

The term in brackets is the survival probability for species 1 in terms of b_2. Maximizing this expression over $0 \le b_2 \le 1$ yields the following optimal conservation expenditures b_2^* and $b_1^* = 1 - b_2^*$:

$$b_1^* = \begin{cases} 1 & ;\ \kappa < 0 \\ \dfrac{1}{1+4\kappa^2} & ;\ \kappa \ge 0 \end{cases} \quad \text{and} \quad b_2^* = \begin{cases} 0 & ;\ \kappa < 0 \\ \dfrac{4\kappa^2}{1+4\kappa^2} & ;\ \kappa \ge 0 \end{cases} \tag{11.22}$$

Figure 11.4 illustrates the result. It shows how the optimal allocation of the conservation budget depends on the interaction strength κ. As long as species 2 has a negative ($\kappa < 0$) or neutral ($\kappa = 0$) influence on the target species 1, the optimal allocation of the conservation budget is to entirely devote it to protection of species 1.[4,5] Obviously, spending money on conserving species 2 which then negatively impacts species 1 will not be optimal if, in the end, all utility derives from species 1. But if the support species 2 has a positive ($\kappa > 0$) influence on the target species 1, it is optimal to allocate a certain fraction of

---

[4]If species 2 has a negative influence on the desired target species and no direct utility in itself, it may even be optimal to not only *not* invest in its protection, but to invest in its reduction. For example, species 2 may be a pest or parasite for species 1 and, for the sake of protecting species 1, it may seem desirable to eliminate this pest or parasite. However, in the formal framework employed here I only consider species *protection* plans, i.e. one can only invest into *enhancing* a species' survival probability.

[5]Note that for vanishing interaction strength, $\kappa = 0$, solution (11.22) reduces to the solution (11.10) obtained in Section 11.4 above for the case of independent species.

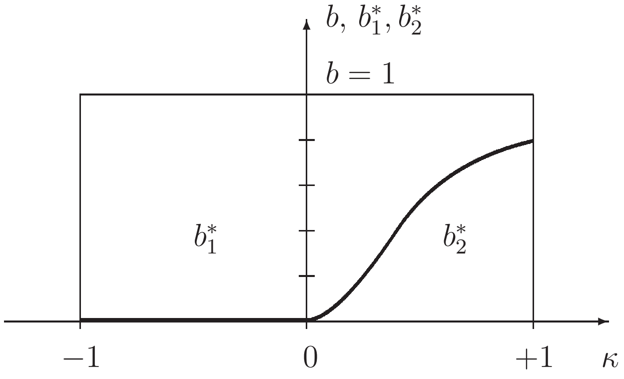

Figure 11.4 Optimal allocation of the conservation budget ($b = 1$) among the target species (b_1^*) and the support species (b_2^*), depending on the interaction strength κ. The curve shows b_2^* as a function of the interaction parameter κ, with the distance between the curve and $b = 1$ corresponding to b_1^*.

the allocation budget to the protection of the support species as well. This fraction grows as the positive interspecific influence (κ) grows in strength.

For $\kappa = +1$ the optimal allocation of the conservation budget is $b_1^* = 0.2$, $b_2^* = 0.8$. In this case the positive influence from species 2 on species 1 is so strong that by spending the largest part of the budget on protecting species 2, one obtains a higher survival probability of species 1 than any direct investment into that species would produce. The reason for this result is in the assumption, illustrated in Figure 11.3, that for $\kappa = +1$ the entire feasible range of conditional survival probabilities for species 1 with species 2 in existence, $[4/5, 1]$, is higher than in the absence of species 2, $[2/5, 3/5]$. As argued above (Figure 11.2d), this corresponds to an evolutionary old ecosystem with a high degree of mutual adaptation among species. Existence of the support species can then provide a better service to the survival of the target species than any direct investment into protecting the target species could possibly achieve. Hence, spending money on increasing the support species' survival probability, thus indirectly also increasing the target species' survival probability, is more cost-effective than spending the entire budget directly on the target species.

The result, thus, is that species interaction can completely reverse the optimal allocation of a conservation budget. In the example studied here, while the entire conservation budget would be allocated to species 1 without any interaction, a strongly positive interaction will make it optimal to allocate almost the entire budget to conservation of species 2.

If one substitutes result (11.22) back into expression (11.21) for the unconditional survival probability of species 1 one obtains

$$p_1^* = \left[\frac{2}{5} + \frac{2}{5}\kappa\sqrt{b_2^*} + \frac{1}{5}\sqrt{1-b_2^*}\right] = \frac{2\sqrt{1+4\kappa^2} + 4\kappa^2 + 1}{5\sqrt{1+4\kappa^2}} . \tag{11.23}$$

Figure 11.5 illustrates this result. It shows how the optimal survival probability

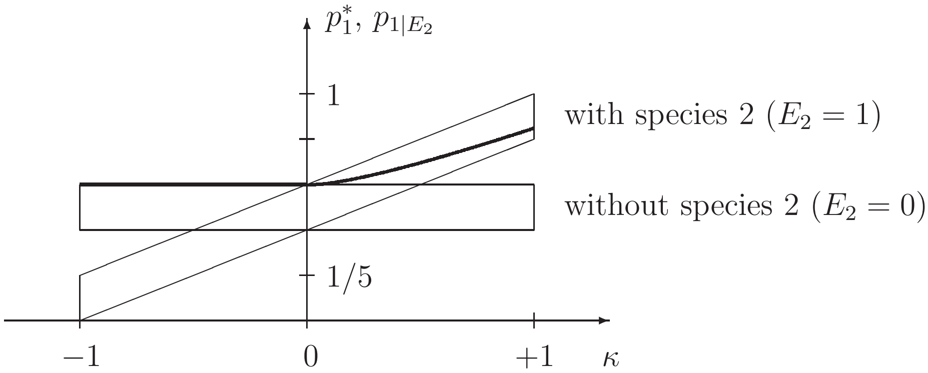

Figure 11.5 Optimal survival probability p_1^* of the target species (thick curve), depending on the interaction strength κ between support and target species.

of the target species, p_1^*, increases with the interaction strength for $\kappa \geq 0$.

11.7 Summary and Discussion of Results

This analysis has shown that taking into account species interactions in an ecosystem is crucial for the optimal allocation of a conservation budget. Compared with policy recommendations obtained under the assumption of independent species, interactions in an ecosystem can reverse the rank ordering of spending priorities among species conservation projects. Hence, an approach to species protection that is efficient in terms of both species conservation and budget resources should be based on a multi-species framework and should take into account the basic underlying ecological relations. Another interesting result is that even if biological conservation decisions are exclusively derived from a utilitarian framework, with species interaction it may be optimal to invest in the protection of species that do not directly contribute to human well-being. This is due to their role for overall ecosystem functioning and for safeguarding the existence of those species that are the ultimate target of environmental policy.

For practical purposes, however, one is confronted with a large extent of uncertainty about the functioning of ecosystems, including fundamental uncertainty about the exact nature of species interaction (Brown et al. 2001, Holling et al. 1995, Loreau et al. 2001, Tilman 1997). In many cases, it is not even known whether two species have a positive or negative interaction. In terms of the model outlined above, this means that neither the exact value of κ nor its sign are known. It may be due to this large ecological uncertainty that the multi-species recovery plans, which have become more and more important in the U.S. Fish and Wildlife Service's approach to protecting endangered species, turned out to be even less successful in terms of species recovery than the more traditional single-species plans (Clark and Harvey 2002).

The description of species interactions proposed here is very simple since it takes into account species and their interaction only on a discrete basis (species i exists/does not exist). A more realistic picture would involve population size and population dynamics for each species. Yet, this would not alter the qualitative results obtained here. The description of species protection plans is equally simple, as it is assumed that each plan affects only the very species at which it is directed. In practice, however, every species protection plan is likely to affect other species in the ecosystem as well.

The analysis here was mainly based on the illustrating example of a two-species-ecosystem with one-way interaction. The absence of feedbacks excludes any kind of complex dynamics among the species. While this is a very simple and special setting, it can be generalized. With n different species, all of which are potentially interacting, there are $n(n-1)$ pairwise directed interactions, leading to indirect interactions among species as well as positive and negative feedback loops. This number rises very fast as n becomes large. Empirical evidence suggests, however, that the vast majority of pairwise interactions in real ecosystems are weak (McCann et al. 1998, Paine 1992, Wootton 1997). The hope may thus be that in applied studies of how to allocate a conservation budget one can safely neglect a whole many interactions, except for the few strong ones for each species, and that there are considerably less than $n(n-1)$ interactions to be taken into account.

However, empirical evidence also suggests that even the weak interactions are important: the complex interdependence of species survival probabilities, together with the existence of extinction thresholds (Lande 1987, Muradian 2001), is known to give rise to so-called extinction cascades (Borrvall et al. 2000, Lundberg et al. 2000). This means that extinction of one species could entail a cascade of further extinctions. Thus, the extinction of some species may threaten even the existence of other species that are only very weakly linked to the former.

11.8 Conclusion: Managing for Ecosystem Health

This discussion suggests that, for conservation purposes, not only are interactions among individual species important, but also the functioning of ecosystems at large is tantamount. Indeed, conservationists have been arguing for years that effective species protection should go beyond targeting individual species, and aim at whole ecosystems or landscapes.[6] This analysis suggests how such a claim can be made more substantial and operational.

In the multi-species-interaction approach taken here, systemic properties of an ecosystem, such as e.g. their structural and functional organization, their productivity, their resilience under disturbances, and their ability to mitigate the impact of various stresses, underly and influence the survival probabilities of individual species. Thus, individual species' survival depends on, and is determined by, what has been called 'ecosystem health'. The concept of ecosystem health is a complex one, as it involves considerations from the natural, social and health sciences (Rapport et al. 1998). Although difficult to measure and operationalize (Mageau et al. 1995), the notion of ecosystem health reminds one that species conservation in situ ultimately depends on certain properties of the entire system in which the target species lives.

As an encompassing and detailed analysis of the myriad of mutual interactions on the species level in an ecosystem may generally not be possible for a particular species protection plan, a useful alternative and complement can be to take a system approach and manage ecosystems for their functions and health.[7] Ecosystem functioning and health is a necessary prerequisite for species conservation in situ.

---

[6]Individual species may nevertheless be of crucial importance for devising, assessing and marketing such a more holistic approach, for instance as so-called 'keystone', 'flagship' or 'umbrella' species (Simberloff 1998).

[7]Mageau et al. (1995), among others, suggest an operational and quantifiable definition of 'ecosystem health' in terms of ecosystem functions.

References

Aaheim, H.A. (1999), 'Climate policy with multiple sources and sinks of greenhouse gases', *Environmental and Resource Economics*, **14**, 413–429.

Aarssen, L.W. (1997), 'High productivity in grassland ecosystems: affected by species diversity or productive species?', *Oikos*, **80**, 183–184.

Acosta, J.J. (2001), *Physische Input-Output-Rechnung: Ansätze, Möglichkeiten und Probleme einer aktivitätsbezogenen Stofflussrechnung auf nationaler und regionaler Ebene*, Karlsruhe: Institut für Regionalwissenschaft, Universität Karlsruhe.

Aczél, J. and Z. Daróczy (1975), *On Measures of Information and their Characterizations*, New York: Academic Press.

Ahmad, Y.J., A. El Serafy and E. Lutz (1989), *Environmental Accounting for Sustainable Development*, Washington DC: World Bank.

Amir, S. (1995), 'Welfare maximization in economic theory: another viewpoint', *Structural Change and Economic Dynamics*, **6**, 359–376.

Amir, S. (1998), 'The role of thermodynamics for ecological economics', *Ecological Economics*, **27**, 213–214.

Anderson, C.L. (1987), 'The production process: inputs and wastes', *Journal of Environmental Economics and Management*, **14**, 1–12.

Ando, A., J. Camm, S. Polasky and A. Solow (1998), 'Species distributions, land values, and efficient conservation', *Science*, **279**, 2126–2128.

Andresen, B., P. Salomon and R.S. Berry (1984), 'Thermodynamics in finite time', *Physics Today*, **37**(9), 62–70.

Arndt, U., W. Nobel and B. Schweizer (1987), *Bioindikatoren: Möglichkeiten, Grenzen und neue Erkenntnisse*, Stuttgart: Ulmer.

Arrow, K.J. (1951), 'Alternative proof of the substitution theorem for Leontief models in the general case', in T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, New York: Wiley, pp. 155–164.

Arrow, K.J. (1965), 'Aspects of the theory of risk-bearing', *Yrjö Jahnssonin Säätio Lecture*, Helsinki, reprinted with modifications in: K.J. Arrow, *Essays in the Theory of Risk Bearing*, Chicago: Markham, 1971.

Arrow, K.J., B. Bolin, R. Costanza, P. Dasgupta, C. Folke, C.S. Holling, B.-O. Jansson, S. Levin, K.-G. Mäler, C. Perrings and D. Pimentel (1995), 'Economic growth, carrying capacity, and the environment', *Science*, **268**, 520–521.

Arrow, K.J. and G. Debreu (1954), 'Existence of an equilibrium for a competitive economy', *Econometrica*, **22**, 265–290.

Arrow, K.J. and A.C. Fisher (1974), 'Preservation, uncertainty and irreversibility', *Quarterly Journal of Economics*, **88**, 312–319.

Arthur, W.B. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal*, **99**, 116–131.

Ayres, R.U. (1978), *Resources, Environment and Economics. Applications of the Materials/Energy Balance Principle*, New York: Wiley.

Ayres, R.U. (1995), 'Thermodynamics and process analysis for future economic scenarios', *Environmental and Resource Economics*, **6**, 207–230.

Ayres, R.U. (1998), 'Eco-thermodynamics: economics and the second law', *Ecological Economics*, **26**, 189–209.

Ayres, R.U. (1999a), 'Materials, economics and the environment', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 867–894.

Ayres, R.U. (1999b), 'The second law, the fourth law, recycling, and limits to growth', *Ecological Economics*, **29**, 473–483.

Ayres, R.U. and L.W. Ayres (eds) (2002), *Handbook of Industrial Ecology*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Ayres, R.U., L.W. Ayres and K. Martinás (1998), 'Exergy, waste accounting, and life-cycle analysis', *Energy*, **23**, 355–363.

Ayres, R.U., L.W. Ayres and B. Warr (2003), 'Exergy, power and work in the US economy 1900–1998', *Energy – The International Journal*, **28**, 219–273.

Ayres, R.U. and A.V. Kneese (1969), 'Production, consumption, and externalities', *American Economic Review*, **59**, 282–297.

Ayres, R.U. and K. Martinás (1995), 'Waste potential entropy: the ultimate ecotoxic?', *Economie Appliquée*, **48**, 95–120.

Ayres, R.U. and U. Simonis (1994), *Industrial Metabolism: Restructuring for Sustainable Development*, Tokyo: United Nations University Press.

Balian, R. (1991), *From Microphysics to Macrophysics*, vol. I, Berlin, Heidelberg and New York: Springer.

Barbier, E.B., J.C. Burgess and C. Folke (1994), *Paradise Lost? The Ecological Economics of Biodiversity*, London: Earthscan.

Barelli, P. and S. de Abreu Pessôa (2003), 'Inada conditions imply that production function must be asymptotically Cobb-Douglas', *Economics Letters*, **81**, 361–363.

Batabyal, A.A. and H. Beladi (2002), 'On the choice of the optimal temporal control in renewable resource management', *Stochastic Environmental Research and Risk Assessment*, **16**, 325–332.

Batabyal, A.A., B. Biswas and B.E. Godfrey (2001), 'On the choice between the stocking rate and time in range management', *Environmental and Resource Economics*, **20**, 211–223.

Bateman, I.J., R.T. Carson, D. Day, M. Hanemann, N. Hanley, T. Hett, M.W. Jones-Lee, G. Loomes, S. Mourato, E. Özdemiroglu, D.W. Pearce, R. Sugden and J. Swanson (2002), *Economic Valuation with Stated Preference Techniques. A Manual*, Cheltenham, UK and Northampton, MA, USA, UK: Edward Elgar.

Baumgärtner, S. (2000), *Ambivalent Joint Production and the Natural Environment. An Economic and Thermodynamic Analysis*, Heidelberg and New York: Physica.

Baumgärtner, S. (2001), 'Thermodynamic and economic notions of efficiency', Contribution to the Symposium *Sustainable Processes and Products*, 30 May 2001, Delft University of Technology, Department of ChemTech.

Baumgärtner, S. (2002), 'Thermodynamics of waste generation', in K. Bisson and J. Proops (eds), *Waste in Ecological Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 13–37.

Baumgärtner, S. (2003), 'Entropy', in International Society for Ecological Economics (ed.), *Online Encyclopedia of Ecological Economics* www.ecoeco.org/publica/encyc.htm.

Baumgärtner, S. (2004a), 'The Inada conditions for material resource inputs reconsidered', *Environmental and Resource Econommics*, **29**(3), 307–322.

Baumgärtner, S. (2004b), 'Measuring the diversity of what? And for what purpose? A conceptual comparison of ecological and economic biodiversity indices', Paper presented at the *Annual Conference of the European Association of Environmental and Resource Economists* (Budapest) and at the *Biennial Conference of the International Society for Ecological Economics* (Montreal).

Baumgärtner, S. (2004c), 'Optimal investment in multi-species protection: interacting species and ecosystem health', *EcoHealth*, **1**(1), 101–110.

Baumgärtner, S. (2004d), 'Price ambivalence of secondary resources: joint production, limits to substitution, and costly disposal', *Resources, Conservation and Recycling*, **43**(1), 95–117.

Baumgärtner, S. (2004e), 'Thermodynamic models: rationale, concepts, and caveats', in J. Proops and P. Safonov (eds), *Modelling in Ecological Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 102–129.

Baumgärtner, S. (2005), 'Temporal and thermodynamic irreversibility in production', *Economic Theory*, **26**(3), 725–728.

Baumgärtner, S. (2006), 'Why the measurement of biodiversity requires prior value judgments', in A. Kontoleon, U. Pascual and T. Swanson (eds), *Frontiers in Biodiversity Economics*, Cambridge: Cambridge University Press.

Baumgärtner, S., C. Becker, M. Faber and R. Manstetten (in press), 'Relative and absolute scarcity of nature. Assessing the roles of economics and ecology for nature conservation', *Ecological Economics*.

Baumgärtner, S. and J. de Swaan Arons (2003), 'Necessity and inefficiency in the generation of waste. A thermodynamic analysis', *Journal of Industrial Ecology*, **7**(2), 113–124.

Baumgärtner, S., H. Dyckhoff, M. Faber, J. Proops and J. Schiller (2001), 'The concept of joint production and ecological economics', *Ecological Economics*, **36**(3), 365–372.

Baumgärtner, S., M. Faber and J. Proops (1996), 'The use of the entropy concept in ecological economics', in M. Faber, R. Manstetten and J. Proops (1996), *Ecological Economics. Concepts and Methods*, Cheltenham, UK and Brookfield, USA: Edward Elgar, pp. 115–135.

Baumgärtner, S., M. Faber and J. Schiller (2006), *Joint Production and Responsibility in Ecological Economics. On the Foundations of Environmental Policy*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Baumgärtner, S. and F. Jöst (2000), 'Joint production, externalities, and the regulation of production networks', *Environmental and Resource Economics*, **16**(2), 229–251.

Baumgärtner, S. and M. Quaas (2005), 'The private and public insurance value of conservative biodiversity management', *Discussion Paper*, Department of Economics, University of Heidelberg.

Baumol, W.J. and W.E. Oates (1988), *The Theory of Environmental Policy*, 2nd edn., Cambridge: Cambridge University Press.

Beard, T.R. and G.A. Lozada (1999), *Economics, Entropy and the Environment. The Extraordinary Economics of Nicholas Georgescu-Roegen*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Becker, C. (2005), 'Wie Ökonomen über Natur denken', in G. Altner, H. Leitschuh-Fecht, G. Michelsen, U.E. Simonis and E.U. von Weizsäcker (eds), *Jahrbuch Ökologie 2006*, München: Beck, pp. 87–98.

Becker, C. and S. Baumgärtner (2005), 'Einleitung', in S. Baumgärtner and C. Becker (eds), *Wissenschaftsphilosophie interdisziplinärer Umweltforschung*, Marburg: Metropolis, pp. 9–23.

Beckerman, W. (1999), 'A pro-growth perspective', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 622–634.

Begon, M. J.L. Harper and C.R. Townsend (1998), *Ecology. Individuals, Populations, and Communities*, 3rd edn., Sunderland: Sinauer.

Behnke, R.H., I. Scoones and C. Kerven (1993), *Range Ecology at Disequilibrium: New Models of Natural Variability and Pastoral Adaptation in African Savannas*, London: Overseas Development Institute.

Bejan, A. (1996), *Entropy Generation Minimization*, Boca Raton: CRC.

Bejan, A. (1997), *Advanced Engineering Thermodynamics*, 2nd edn., New York: Wiley.

Bejan, A., G. Tsatsaronis and M. Moran (1996), *Thermal Design and Optimization*, New York: Wiley.

Benhabib, J. and K. Nishimura (1979), 'The Hopf bifurcation and the existence and stability of closed orbits in multisector models of optimal economic growth', *Journal of Economic Theory*, **21**, 421–444.

Berger, W.H. and F.L. Parker (1970), 'Diversity of planktonic Foraminifera in deep sea sediments', *Science*, **168**, 1345–1347.

Berry, R.S. and B. Andresen (1982), 'Thermodynamic constraints in economic analysis', in W.C. Schieve and P. Allen (eds), *Self-Organization and Dissipative Structures. Applications in the Physical and Social Sciences*, Austin: University of Texas Press, pp. 323–338.

Berry, R.S., P. Salamon and G.M. Heal (1978), 'On a relation between economic and thermodynamic optima', *Resources and Energy*, **1**, 125–137.

Besley, T. (1995), 'Savings, credit and insurance', in H.B. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics*, vol. III, Amsterdam: North-Holland, pp. 2123–2207.

Beukes, P.C., R.M. Cowling and S.I. Higgins (2002), 'An ecological economic simulation model of a non-selective grazing system in the Nama Karoo, South Africa', *Ecological Economics*, **42**, 221–242.

Biervert, B. and M. Held (eds) (1994), *Das Naturverständnis der Ökonomik. Beiträge zur Ethikdebatte in den Wirtschaftswissenschaften.* Frankfurt: Campus.

Binswanger, H.C. (1991), 'Brazilian policies that encourage the deforestation in the Amazon', *World Development*, **19**, 821.

Birol, E., A. Kontoleon and M. Smale (2005a), 'Using a choice experiment to estimate the demand of Hungarian farmers for food security and agrobiodiversity during economic transition', Paper presented at the 7th BIOECON conference, 20–21 September 2005, Cambridge, UK.

Birol, E., M. Smale and A. Gyovoi (2005b), 'Farmer management of agricultural biodiversity in Hungary's transition economy', in M. Smale (ed.), *Valuing Crop Biodiversity: On-farm Genetic Resources and Economic Change*, Wallingford, UK: CAB International.

Bisson, K. and J. Proops (eds) (2002), *Waste in Ecological Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Blaug, M. (1996), *Economic Theory in Retrospect*, 5th edn., Cambridge: Cambridge University Press.

Bohn, H. and R.T. Deacon (2000), 'Ownership risk, investment, and the use of natural resources', *American Economic Review*, **90**(3), 526–549.

Boltzmann, L. (1877), 'Über die Beziehung eines allgemeinen mechanischen Satzes zum zweiten Hauptsatz der Wärmetheorie', *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien*, Abt. 2, **75**, 67–73.

Bond, W.J. (1993), 'Keystone species', in E.-D. Schulze and H.A. Mooney (eds), *Biodiversity and Ecosystem Functions*, Berlin, Heidelberg and New York: Springer, pp. 238–253.

Borrvall, C., B. Ebenman, and T. Jonsson (2000), 'Biodiversity lessens the risk of cascading extinction in model food webs', *Ecology Letters*, **3**(2), 131–136.

Borrvall, C., B. Ebenman, and T. Jonsson (2000), 'Biodiversity lessens the risk of cascading extinction in model food webs', *Ecology Letters*, **3**(2), 131–136.

Boulding, K.E. (1966), 'The economics of the coming spaceship Earth', in H. Jarrett (ed.), *Environmental Quality in a Growing Economy*, Baltimore, MD, USA and London, UK: Johns Hopkins University Press, pp. 3–14.

Bowles, M.L. and C.J. Whelan (eds) (1994), *Restoration of Endangered Species*, Cambridge: Cambridge University Press.

Briske, D.D., S.D. Fuhlendorf and F.E. Smeins (2003), 'Vegetation dynamics on rangelands: A critique of the current paradigms', *Journal of Applied Ecology*, **40**, 601–614.

Brodyansky, V.M., M.V. Sorin and P. Le Goff (eds) (1994), *The Efficiency of Industrial Processes: Exergy Analysis and Optimization*, Amsterdam: Elsevier.

Brown, G.M., Jr. and J.F. Shogren (1998), 'Economics of the Endangered Species Act', *Journal of Economic Perspectives*, **12**(3), 3–20.

Brown, J.H. T.G. Whitham, S.K. Morgan Ernest and C.A. Gehring (2001), 'Complex species interactions and the dynamics of ecological systems: long-term experiments', *Science*, **293**, 643–650.

Brown, L.R. (2001), *Eco-Economy. Building an Economy for the Earth*, New York: Norton.

Buenstorf, G. (2004), *The Economics of Energy and the Production Process: An Evolutionary Approach*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Burgman, M.A., S. Ferson and H.R. Akçakaya (1993), *Risk Assessment in Conservation Biology*, London: Chapman & Hall.

Burley, P. and J. Foster (eds) (1994), *Economics and Thermodynamics: New Perspectives on Economic Analysis*, Dordrecht: Kluwer.

Burness, H.S., R.G. Cummings, G. Morris and I. Paik (1980), 'Thermodynamic and economic concepts as related to resource-use policies', *Land Economics*, **56**, 1–9.

Caillaud, B., G. Dionne and B. Jullien (2000), 'Corporate insurance with optimal financing contracts', *Economic Theory*, **16**, 77–105.

Callen, H.B. (1985), *Thermodynamics and an Introduction to Thermostatics*, 2nd edn., New York: Wiley.

Candeal, J., J. Miguel, E. Indurain and G. Mehta (2001a), 'On a theorem of Cooper', *Journal of Mathematical Analysis and Applications*, **258**, 701–710.

Candeal, J., J. Miguel, E. Indurain and G. Mehta (2001b), 'Utility and entropy', *Economic Theory*, **17**, 233–238.

Casperson, J.P. and S.W. Pacala (2001), 'Successional diversity and forest ecosystem function', *Ecological Research*, **16**, 895–904.

[CBD] *Convention on Biological Diversity*, United Nations Conference on Environment and Development, Rio de Janeiro, Brazil, 1992.

Ceballos, G. and P.R. Ehrlich (2002), 'Mammal population losses and the extinction crisis', *Science*, **296**, 904–907.

Chambers, R.G. (1989), 'Insurability and moral hazard in agricultural insurance markets', *American Journal of Agricultural Economics*, **71**, 604–616.

Chapin, F.S. III and G.R. Shaver (1985), 'Individualistic growth response of tundra plant species to environmental manipulations in the field', *Ecology*, **66**, 564–576.

Chesson, P.L. (2000), 'Mechanisms of maintenance of species diversity', *Annual Review of Ecology and Systematics*, **31**, 343–366.

Chesson, P., S. Pacala and C. Neuhauser (2002), 'Environmental niches and ecosystem functioning',in A. Kinzig, D. Tilman and S. Pacala (eds), *Functional Consequences of Biodiversity: Experimental Progress and Theoretical Extensions*, Princeton, NJ, USA: Princeton University Press, pp. 213–245.

Clark, J.A. and E. Harvey (2002), 'Assessing multi-species recovery plans under the Endangered Species Act', *Ecological Applications*, **12**(3), 655–662.

Clausius, R. (1854), *Fortschritte der Physik*, **10**.

Clausius, R. (1865), 'Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie', *Annalen der Physik*, **125**, 353–400.

Cleveland, C.J. and M. Ruth (1997), 'When, where and by how much does thermodynamics constrain economic processes? A survey of Nicholas Georgescu-Roegen's contribution to ecological economics', *Ecological Economics*, **22**, 203–223.

Ciriacy-Wantrup, S.V. (1965), 'A safe minimum standard as an objective of conservation policy', in I. Burton and R. Kates (eds), *Readings in Resource Management and Conservation*, Chicago: University of Chicago Press, pp. 45–70.

Connelly, L. and K.P. Koshland (2001), 'Exergy and industrial ecology', *Exergy – An International Journal*, **1**, 146–165 and 234–255.

Cooper, J. (1967), 'The foundations of thermodynamics', *Journal of Mathematical Analysis and Applications*, **17**, 172–193.

Cornelissen, R.L. and G.G. Hirs (1999), 'Exergy analysis in the process industry', in A. Bejan and E. Mamut (eds), *Thermodynamic Optimization of Complex Energy Systems*, Dordrecht: Kluwer.

Cornelissen, R.L., P.A. Nimwegen, and G.G. Hirs (2000), 'Exergetic Life-Cycle analysis', in *Proceedings of ECOS 2000*, Enschede.

Costanza, R. (1981), 'Embodied energy, energy analysis, and economics', in H.E. Daly (ed.), *Energy, Economics, and the Environment*, Boulder: Westview.

Costanza, R. (1989), 'What is ecological economics?', *Ecological Economics*, **1**(1), 1–7.

Costanza, R. (ed.) (1991), *Ecological Economics. The Science and Managament of Sustainability*, New York: Columbia University Press.

Costanza, R., J. Cumberland, H. Daly, R. Goodland and R. Norgaard (1997a), *An Introduction to Ecological Economics*, Boca Raton, FL: St. Lucia.

Costanza, R., R. d'Arge, R. de Groot, S. Farber, W. Grasso, B. Hannon, K. Limburg, S. Naeem, R.V. O'Neill, J. Paruelo, R.G. Raskin, P. Sutton and M. van den Belt (1997b), 'The value of the world's ecosystem services and natural capital', *Nature*, **387**, 253–260.

Costanza, R., C. Perrings ad C.J. Cleveland (eds) (1997c), *The Development of Ecological Economics*, Cheltenham, UK and Lyme, USA: Edward Elgar.

Costello, C. and S. Polasky (2004), 'Dynamic reserve site selection', *Resources and Energy Economics*, **26**, 157–174

Cottingham, K.C., B.L. Brown, and J.T. Lennon (2001) 'Biodiversity may regulate the temporal variability of ecological systems', *Ecology Letters*, **4**, 72–85.

Cowling, R.M. (2000), 'Challenges to the "new" rangeland science', *Trends in Ecology & Evolution*, **15**, 303–304.

Cullen, R., G.A. Fairburn and K.F.D. Hughey (2001), 'Measuring the productivity of threatened-species programs', *Ecological Economics*, **39**, 53–66.

Craig, P.P. (2001), 'Energy limits to recycling', *Ecological Economics*, **36**, 373–384.

Creyts, J.C. (2000), *Use of Extended Exergy Analysis as a Tool to Optimize the Environmental Performance of Industrial Processes*, Ph.D. Thesis, Berkeley: University of California, Department of Mechanical Engineering.

Crocker, T.D. and J.F. Shogren (1999), 'Endogenous environmental risk', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 215–222.

Crocker, T.D. and J.F. Shogren (2001), 'Ecosystems as lotteries', in H. Folmer, H.L. Gabel, S. Gerking and A. Rose (eds), *Frontiers of Environmental Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 250–271.

Crocker, T.D. and J.F. Shogren (2003), 'Choosing environmental risks', in H. Folmer and T. Tietenberg (eds), *The International Yearbook of Environmental and Resource Economics 2003/04*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 36–81.

Crozier, R.H. (1992), 'Genetic diversity and the agony of choice', *Biological Conservation*, **61**, 11–15.

Cullen, R., G.A. Fairburn and K.F.D. Hughey (2001), 'Measuring the productivity of threatened-species programs', *Ecological Economics*, **39**, 53–66.

Daily, G.C. (1997a), 'Introduction: What are ecosystem services?', in G.C. Daily (ed.), *Nature's Services: Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press, pp. 1–10.

Daily, G.C. (ed.) (1997b), *Nature's Services: Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press.

Daly, H.E. (1973), *Toward a Steady-State Economy*, San Francisco: Freeman.

Daly, H.E. (1991)[1977], *Steady State Economics. The Economics of Biophysical Equilibrium and Moral Growth*, 2nd ed. (1st ed. 1977), San Francisco: Freeman.

Daly, H.E. (1992a), 'Allocation, distribution, and scale: towards an economics that is efficient, just, and sustainable', *Ecological Economics*, **6**, 185–193.

Daly, H.E. (1992b), 'Is the entropy law relevant to the economics of natural resources? Yes, of course it is!', *Journal of Environmental Economics and Management*, **23**, 91–95.

Daly, H.E. (1996), *Beyond Growth. The Economics of Sustainable Development*, Boston: Beacon.

Daly, H.E. (ed.) (1997), *The Contribution of Nicholas Georgescu Roegen*, Special issue of *Ecological Economics*, **22**(3).

Daly, H.E. (1999), 'Steady-state economics: avoiding uneconomic growth', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 635–642.

Daly, H.E. (1997b), 'Georgescu-Roegen versus Solow/Stiglitz' *Ecological Economics*, **22**, 261–266.

Daly, H.E. (1999), 'Steady-state economics: avoiding uneconomic growth', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 635–642.

Dasgupta, P. (1993), *An Inquiry into Well-Being and Destitution*, Oxford: Clarendon.

Dasgupta, P. (1995), 'Population, poverty, and the local environment', *Scientific American*, **272**, 40ff.

Dasgupta, P. (2000), 'Valuing biodiversity', in S.A. Levin (ed.), *Encyclopedia of Biodiversity*, New York: Academic Press.

Dasgupta, P. and G. Heal (1974), 'The optimal depletion of exhaustible resources', *Review of Economic Studies*, **41** *(Symposium on the Economics of Exhaustible Resources)*, 3–28.

Dasgupta, P. and G. Heal (1979), *Economic Theory and Exhaustible Resources*, Cambridge, UK: Cambridge University Press.

Dawson, D. and J.F. Shogren (2001), 'An update on priorities and expenditures under the Endangered Species Act', *Land Economics*, **77**(4), 527–532.

de Bruyn, S. (1997), 'Explaining the environmental Kuznets curve: structural change and international agreements in reducing sulphur emissions', *Environment and Development Economics*, **2**, 485–503.

de Ruiter, P.C., B. Griffiths and J.C. Moore (2002), 'Biodiversity and stability in soil ecosystems: patterns, processes and the effects of disturbance', in M. Loreau, S. Naeem and P. Inchausti (eds), *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives*, Oxford: Oxford University Press, pp. 102–11.

de Swaan Arons, J. and H.J. van der Kooi (2001), 'Towards a metabolic society: a thermodynamic view', *Green Chemistry*, **3**(4), G53–G55.

de Swaan Arons, J., H.J. van der Kooi and K. Sankaranarayanan (2004), *Efficiency and Sustainability in the Energy and Chemical Industries*, New York: Dekker.

Dean, W.R.J. and I.A.W. MacDonald (1994), 'Historical changes in stocking rates of domestic livestock as a measure of semi-arid and arid rangeland degradation in the Cape Province, South Africa', *Journal of Arid Environments*, **26**, 281–298.

Debreu, G. (1959), *Theory of Value. An Axiomatic Study of Economic Equilibrium*, New York: Wiley.

Debreu, G. (1962), 'New concepts and techniques for equilibrium analysis', *International Economic Review*, **3**, 257–273.

Dewulf, J., J.M. Mulder, M.M.D. van den Berg, H. Van Langenhove, H.J. van der Kooi and J. de Swaan Arons (2000), 'Illustrations towards quantifying the sustainability of technology', *Green Chemistry*, **2**(3), 108–114.

Di Falco, S., J.-P. Chavas and M. Smale (2005), 'Farmer management of production risk on degraded lands: wheat diversity in Tigray region, Ethiopia', Paper presented at the 7th BIOECON Conference, 20–21 September 2005, Cambridge, UK.

Di Falco, S. and C. Perrings (2003), 'Crop genetic diversity, productivity and stability of agroecosystems: a theoretical and empirical investigation', *Scottish Journal of Political Economy*, **50**(2), 207–216.

Di Falco, S. and C. Perrings (2005), 'Crop biodiversity, risk management and the implications of agricultural assistance', *Ecological Economics*, **55**(4), 459–466.

Diehl, M., D.B. Leineweber and A.A.S. Schäfer (2001), 'MUSCOD-II users' manual', *Preprint 2001-25*, Interdisciplinary Center for Scientific Computing, University of Heidelberg.

Dixit, A.K. (1992), 'Investment and hysteresis', *Journal of Economic Perspectives*, **6**(4), 107–132.

Dixit, A.K., and R.S. Pindyck (1994), *Investment under Uncertainty*, Princeton: Princeton University Press.

Doak, D.F., D. Bigger, E.K. Harding, M.A. Marvier, R.E. O'Malley and D. Thomson (1998), 'The statistical inevitability of stability-diversity relationships in community ecology', *The American Naturalist*, **151**, 264–276.

Dosi, G. and J.S. Metcalfe (1991), 'On some notions of irreversibility in economics', in P.P. Saviotti and J.S. Metcalfe (eds), *Evolutionary Theories of Economic and Technological Change. Present Status and Future Prospects*, Chur: Harwood, pp. 133–159.

Dyckhoff, H. (1983), 'Inada-Bedingungen und Eigenschaften der neoklassischen Produktionsfunktion', *Zeitschrift für die gesamte Staatswissenschaft*, **139**(1), 146–154.

Dyckhoff, H. (1994), *Betriebliche Produktion. Theoretische Grundlagen einer umweltorientierten Produktionswirtschaft*, 2nd edn., Berlin, Heidelberg and New York: Springer.

Eatwell, J., M. Milgate and R. Newman (eds) (1987), *The New Palgrave. A Dictionary of Economics*, London: Macmillan.

Ehrlich, J. and G.S. Becker (1972), 'Market insurance, self-insurance and self-protection', *Journal of Political Economy*, **80**, 623–648.

Ehrlich, P.R. and A.H. Ehrlich (1981), *Extinction: The Causes and Consequences of the Disappearance of Species*, New York: Random House.

Ehrlich, P.R. and A.H. Ehrlich (1992), 'The value of biodiversity', *Ambio*, **21**, 219–226.

Ehrlich, P.R. and J.P. Holdren (1971), 'Impact of population growth', *Science*, **171**, 1212–1217.

Elton, C.S. (1927), *Animal Ecology*, New York: Macmillan.

Elton, C.S. (1958), *Ecology of Invasions by Animals and Plants*, London: Methuen.

Emmerson, M.C., M. Solan, C. Emes, D.M. Peterson and D. Raffaelli (2001), 'Consistent patterns and the idiosyncratic effects of biodiversity in marine ecosystems', *Nature*, **411**, 73–77.

Ernest, S.K.M. and J.H. Brown (2001), 'Delayed compensation for missing keystone species by colonization', *Science*, **292**, 101–104.

Erwin, T.L. (1991), 'An evolutionary basis for conservation strategies', *Science*, **253**, 750–752.

Ethridge, D. (1973), 'The inclusion of wastes in the theory of the firm', *Journal of Political Economy*, **81**, 1430–1441.

Ewel, J.J. (1986), 'Designing agroecosystems for the humid tropics', *Annual Review of Ecology and Systematics*, **17**, 245–271.

Faber, M. (1985), 'A biophysical approach to the economy: entropy, environment and resources', in W. van Gool and J.J.C. Bruggink (eds), *Energy and Time in the Economic and Physical Sciences*, Amsterdam: North-Holland, pp. 315–335.

Faber, M., F. Jöst and R. Manstetten (1995), 'Limits and perspectives on the concept of sustainable development', *Economie Appliquée*, **48**, 233–251.

Faber, M., R. Manstetten and J. Proops (1996), *Ecological Economics. Concepts and Methods*, Cheltenham, UK and Brookfield, USA: Edward Elgar.

Faber, M., H. Niemes and G. Stephan (1995)[1983], *Entropy, Environment and Resources: An Essay in Physico-Economics*, 2nd edn., 1st edn. 1983 in German, Berlin, Heidelberg and New York: Springer.

Faber, M. and J. Proops (1985), 'Interdisciplinary research between economists and physical scientists: retrospect and prospect', *Kyklos*, **38**, 599–616.

Faber, M. and J. Proops (1998), *Evolution, Time, Production and the Environment*, 3rd rev. edn., Berlin, Heidelberg and New York: Springer.

Faber, M., J. Proops and S. Baumgärtner (1998), 'All production is joint production. A thermodynamic analysis', in S. Faucheux, J. Gowdy and I. Nicolaï (eds), *Sustainability and Firms. Technological Change and the Changing Regulatory Environment*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 131–158.

Falk, I. and R. Mendelsohn (1993), 'The economics of controlling stock pollutants: an efficient strategy for greenhouse gases', *Journal of Environmental Economics and Pollution*, **25**, 76–88.

Falk, G. and W. Ruppel (1976), *Energie und Entropie. Eine Einführung in die Thermodynamik*, Berlin, Heidelberg and New York: Springer.

Färe, R. and D. Primont (2002), 'Inada conditions and the law of diminishing returns', *International Journal of Business and Economics*, **1**(1), 1–8.

Farnsworth, N.R. (1988), 'Screening plants for new medecines', in E.O. Wilson (ed.), *BioDiversity*, Washington, DC: National Academy Press, pp. 83–97.

Feichtinger, G., A. Novak and F. Wirl (1994), 'Limit cycles in intertemporal adjustment models: Theory and application', *Journal of Economic Dynamics and Control*, **18**, 353–380.

Field, C.B. (1995), 'Productive capacity and biomass accumulation', in United Nations Environment Programme (UNEP) (ed.), *Global Biodiversity Assessment*, Cambridge, UK: Cambridge University Press, pp. 402–406.

Fisher, A.C. (2000), 'Editorial: Introduction to special issue on irreversibility', *Resource and Energy Economics*, **22**, 189–196.

Fisher, A.C. and W.M. Hanemann (1986), 'Option value and the extinction of species', in V.K. Smith (ed.), *Advances in Applied Microeconomics*, JAI, Greenwich, pp. 169–190.

Forster, B.A. (1973), 'Optimal capital accumulation in a polluted environment', *Southern Economic Journal*, **39**, 544–547.

Freeman III, A.M. (2003), *The Measurement of Environmental and Resource Values. Theory and Methods*, 2nd edn., Washington, DC: Resources for the Future.

Fridley, J.D. (2001), 'The influence of species diversity on ecosystem productivity: how, where and why?', *Oikos*, **93**, 514–526.

Fridley, J.D. (2003), 'Diversity effects on production in different light and fertility environments: an experiment with communities of annual plants' *Journal of Ecology*, **91**, 396– 406.

Fromm, O. (2000), 'Ecological structure and functions of biodiversity as elements of its total economic value', *Environmental and Resource Economics*, **16**, 303–328.

Frost, T.M., S.R. Carpenter, A.R. Ives and T.K. Kratz (1995), 'Species compensation and complementarity in ecosystem function', in C.G. Jones and J.H. Lawton (eds), *Linking Species and Ecosystems*, San Diego, CA, USA: Chapman & Hall, pp. 224–239.

Fuhlendorf, S.D. and D.M. Engle (2001), 'Restoring heterogeneity on rangelands: Ecosystem management based on evolutionary grazing patterns', *BioScience*, **51**, 625–632.

Garnier, E., M.-L. Navas, M.P. Austin, J.M. Lilley and R.M. Gifford (1997), 'A problem for biodiversity–productivity studies: how to compare the productivity of multispecific plant mixtures to that of monocultures?', *Acta Oecologica*, **18**, 657–670.

Gaston, K.J. (1996), *Biodiversity. A Biology of Numbers and Difference*, Oxford, UK: Blackwell.

Gaston, K.J. (2000), 'Global patterns in biodiversity', *Nature*, **405**, 220–227.

Geisendorf, S., S. Gronemann and U. Hampicke (1998), *Die Bedeutung des Naturvermögens und der Biodiversität für eine nachhaltige Wirtschaftsweise. Möglichkeiten und Grenzen ihrer Erfaßbarkeit und Wertmessung*, Berlin Erich Schmidt.

Georgescu-Roegen, N. (1951), 'Some properties of a generalized Leontief model', in T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, New York: Wiley, pp. 165–73.

Georgescu-Roegen, N. (1971), *The Entropy Law and the Economic Process*, Cambridge, MA: Harvard University Press.

Georgescu-Roegen, N. (1975), 'Energy and economic myths', *Southern Economic Journal*, **41**, 347–81.

Georgescu-Roegen, N. (1979), 'Energy analysis and economic valuation', *Southern Economic Journa*l, **45**, 1023–1058.

Giess, W. (1998), 'A preliminary vegetation map of Namibia (3rd rev. ed.)', *Dinteria*, **4**, 5–114.

Givnish, T.J. (1994), 'Does diversity beget stability?', *Nature*, **371**, 113–114.

Gollier, C. (2001), *The Economics of Risk and Time*, Cambridge, MA: MIT Press.

Good, I.J. (1953), 'The population frequencies of species and the estimation of population parameters', *Biometrika*, **40**, 237–264.

Goulder, L.H. and D. Kennedy (1997), 'Valuing ecosystem services: philosophical bases and empirical methods', in G.C. Daily (ed.), *Nature's Services. Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press, pp. 23–47.

Gradus, R. and S. Smulders (1993), 'The trade-off between environmental care and long-term growth', *Journal of Economics*, **58**, 25–51.

Gradus, R. and S. Smulders (1996), 'Pollution abatement and long-term growth', *European Journal of Political Economy*, **12**, 506–532.

Gray, L.C. (1913), 'The economic possibilities of conservation', *Quarterly Journal of Economics*, **27**, 497–519.

Gray, L.C. (1914), 'Rent under the assumption of exhaustibility', *Quarterly Journal of Economics*, **28**, 466–489.

Griffiths, B.S., K. Ritz, R.D. Bardgett, R. Cok, S. Christensen, F. Ekelund, S.J. Sorenson, E. Baath, J. Bloem, P.C. De Ruiter, J. Dolfing and B. Nicolardot (2000), 'Ecosystem response of pasture soil communities to fumigationinduced microbial diversity reductions: an examination of the biodiversity-ecosystem function relationship', *Oikos*, **90**, 279–294.

Grime, J.P. (1997), 'Biodiversity and ecosystem function: the debate deepens', *Science*, **277**, 1260–1261.

Grime, J.P., V.K. Brown, K. Thompson, G.J. Masters, S.H. Hillier, I.P. Clarke, A.P. Askew, D. Corker and P. Kielty (2000), 'The response of two contrasting limestone grasslands to simulated climate change', *Science*, **289**, 762–765.

Grimm, V. and C. Wissel (1997), 'Babel, or the ecological stability discussions: An inventory and analysis of terminology and a guide for avoiding confusion', *Oecologia*, **109**, 323–334.

Groombridge, B. (ed.) (1992), *Global Biodiversity: Status of the World's Living Resources. A Report Compiled by the World Conservation Monitoring Centre*, London: Chapman & Hall.

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

Hall, S.J.G. (1992), 'Animal use', in B. Groombridge (ed.), *Global Biodiversity: Status of the World's Living Resources. A Report Compiled by the World Conservation Monitoring Centre*, London: Chapman & Hall, pp. 359–406.

Hampicke, U. (1991), *Naturschutzökonomie*, Stuttgart: Ulmer.

Hampicke, U. (1993), 'Möglichkeiten und Grenzen der monetären Bewertung von Natur', in H. Schnabel (ed.), *Ökointegrative Gesamtrechnung*, Berlin: De Gruyter.

Hanemann, W.M. (1989), 'Information and the concept of option value', *Journal of Environmental Economics and Management*, **16**, 23–27.

Hanley, N. and C.L. Spash (1993), *Cost-Benefit Analysis and the Environment*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Hanley, N., J.F. Shogren and B. White (1997), *Environmental Economics in Theory and Practice*, London: Macmillan.

Hannon, B. (1973), 'An energy standard of value', *American Academy of Political Science Annual*, **410**, 139–153.

Hannon, B. (1979), 'Total energy cost in ecosystems', *Journal of Theoretical Biology*, **56**, 279–293.

Hannon, B., R. Costanza and R.A. Herendeen (1986), 'Measures of energy cost and value in ecosystems', *Journal of Environmental Economics and Management*, **13**, 391–401.

Hardin, G. (1968), 'The tragedy of the commons', *Science*, **162**, 1243–1248.

Harper, J.L. (1977), *Population Biology of Plants*, London and New York: Academic Press.

Hardy, C. and T.E. Graedel (2002), 'Industrial ecosystems as foodwebs', *Journal of Industrial Ecology*, **6**(1), 29–38.

Hartewick, J. and N. Olewiler (1998), *The Economics of Natural Resource Use*, Boston: Addison-Wesley.

Heady, H.F. (1970), 'Grazing systems: terms and definitions', *Journal of Range Management*, **23**, 59–61.

Heady, H.F. (1999), 'Perspectives on rangeland ecology and management', *Rangelands*, **21**, 23–33.

Heal, G. (2000), 'Biodiversity as a commodity', in S.A. Levin (ed.), *Encyclopedia of Biodiversity*, vol. 1, New York: Academic Press, pp. 359–376.

Heal, G., B. Walker, S. Levin, K. Arrow, P. Dasgupta, G. Daily, P. Ehrlich, K.-G. Mäler, N. Kautsky, J. Lubchenco, S. Schneider and D. Starrett (2004), 'Genetic diversity and interdependent crop choices in agriculture', *Resource and Energy Economics*, **26**, 175–184.

Hector, A., E. Bazeley-White, M. Loreau, S. Otway and B. Schmid, (2002), 'Overyielding in plant communities: testing the sampling effect hypothesis with replicated biodiversity experiments', *Ecology Letters*, **5**, 502–511.

Hector, A. B. Schmid, C. Beierkuhnlein, M.C. Caldeira, M. Diemer, P.G. Dimitrakopoulos, J.A. Finn, H. Freitas, P.S. Giller, J. Good, R. Harris, K. Huss-Danell, J. Joshi, A. Jumpponen, C. Korner, P. Leadley, M. Loreau, A. Minns, C. Mulder, G. O'Donovan, S. Otway, J. Pereira, A. Prinz, D. Read, M. Scherer-Lorenzen, E. Schulte, A. Siamantziouras, E. Spehn, A. Terry, A. Troumbis, F. Woodward, S. Yachi and J. Lawton (1999), 'Plant diversity and productivity experiments in European grasslands', *Science*, **286**, 1123–1127

Hector, A., B. Schmid, C. Beierkuhnlein, M.C. Caldeira, M. Diemer, P.G. Dimitrakopoulos, J.A. Finn, H. Freitas, P.S. Giller, J. Good, R. Harris, K. Huss-Danell, J. Joshi, A. Jumpponen, C. Korner, P. Leadley, M. Loreau, A. Minns, C. Mulder, G. O'Donovan, S. Otway, J. Pereira, A. Prinz, D. Read, M. Scherer-Lorenzen, E. Schulte, A. Siamantziouras, E. Spehn, A. Terry, A. Troumbis, F. Woodward, S. Yachi and J. Lawton (2000), 'No consistent effect of plant diversity on productivity? Response', *Science*, **289**, 1255a.

Hein, L. and H.-P. Weikard (2004), 'Managing stochastic dynamic ecosystems: livestock grazing in a Sahelian rangeland', Paper presented at the Annual Conference of the *European Association of Environmental and Resource Economists*, Budapest, June 2004.

Heintz, A. and G.A. Reinhard (1993), *Chemie und Umwelt*, 3., neubearbeitete Auflage, Braunschweig und Wiesbaden: Vieweg.

Henry, C. (1974), 'Investment decisions under uncertainty: the irreversibility effect', *American Economic Review*, **64**, 1006–1012.

Hill, B. (1997), *Innovationsquelle Natur*, Aachen: Shaker.

Hill, M.O. (1973), 'Diversity and evenness: a unifying notation and its consequences', *Ecology*, **54**, 427–431.

Hinderink, P., H.J. van der Kooi and J. de Swaan Arons (1999), 'On the efficiency and sustainability of the process industry', *Green Chemistry*, **1**(6), G176–G180.

Hoekstra, J.M., J.A. Clark, P.D. Boersma and W.F. Fagan (2002), 'A comprehensive review of Endangered Species Act recovery plans', *Ecological Applications*, **12**(3), 630–640.

Hoel, M. (1978), 'Resource extraction and recycling with environmental costs', *Journal of Environmental Economics and Management*, **5**, 220–235.

Holdren, J.P. and P.R. Ehrlich (1974), 'Human population and the global environment', *American Scientist*, **62**, 282–292.

Holling, C.S. (1986), 'Resilience of ecosystems: local surprise and global change', in W.C. Clark and R.E. Munn (eds), *Sustainable Development of the Biosphere*, Cambridge, UK: Cambridge University Press, pp. 292–317.

Holling, C.S., D.W. Schindler, B.W. Walker and J. Roughgarden (1995), 'Biodiversity in the functioning of ecosystems: an ecological synthesis', in C. Perrings, K.-G. Mäler, C. Folke, C.S. Holling and B.-O. Jansson (eds), *Biodiversity Loss. Economic and Ecological Issues*, Cambridge: Cambridge University Press, pp. 44–83.

Hooper, D.U., F.S. Chapin III, J.J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J.H. Lawton, D.M. Lodge, M. Loreau, S. Naeem, B. Schmid, H. Setälä, A.J. Symstad, J. Vandermeer and D.A. Wardle (2005), 'Effects of biodiversity on ecosystem functioning: a consensus of current knowledge', *Ecological Monographs*, **75**(1), 3–35.

Hooper, D., D. Hawksworth and S. Dhillion (1995), 'Microbial diversity and ecosystem processes', in United Nations Environment Programme (UNEP) (ed.), *Global Biodiversity Assessment*, Cambridge, UK: Cambridge University Press, pp. 433—443.

Hooper, D.U., M. Solan, A. Symstad, S. Díaz, M.O. Gessner, N. Buchmann, V. Degrange, P. Grime, F. Hulot, F. Mermillod-Blondin, J. Roy, E. Spehn and L. van Peer (2002), 'Species diversity, functional diversity and ecosystem functioning', in M. Loreau, S. Naeem and P. Inchausti (eds), *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives*, Oxford, UK: Oxford University Press, pp. 195-–208.

Horowitz, J.K. and E. Lichtenberg (1993), 'Insurance, moral hazard and chemical use in agriculture', *American Journal of Agricultural Economics*, **75**, 926–935.

Horowitz, J.K. and E. Lichtenberg (1994a), 'Crop insurance and agricultural chemical use', in D. Hueth and H. Furtan (eds), *Economics of Agricultural Crop Insurance: Theory and Evidence*, Norwell, MA: Kluwer.

Horowitz, J.K. and E. Lichtenberg (1994b), 'Risk-increasing and risk-reducing effects of pesticides', *Journal of Agricultural Economics*, **45**(1), 82–89.

Hotelling, H. (1931), 'The economics of exhaustible resources', *Journal of Political Economy*, **39**, 137–175.

Huang, K. (1987), *Statistical Mechanics*, 2nd edn., New York: Wiley.

Hueting, R. et al. (1998), 'The concept of environmental functions and its valuation', *Ecological Economics*, **25**, 31–35.

Hughes, J.B. and J. Roughgarden (1998), 'Aggregate community properties and the strength of species' interactions', *Proceeding of the National Academy of Sciences USA*, **95**, 6837–6842.

Hurlbert, S.H. (1971), 'The non-concept of species diversity: a critique and alternative parameters', *Ecology*, **52**, 577–586.

Huston, M.A. (1997), 'Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity', *Oecologia*, **110**, 449–460.

Huston, M.A., L.W. Aarssen, M.P. Austin, B.S. Cade, J.D. Fridley, E. Garnier, J.P. Grime, J. Hodgson, W.K. Lauenroth, K. Thompson, J.H. Vandermeer and D.A. Wardle (2000), 'No consistent effect of plant diversity on productivity', *Science*, **289**, 1255.

Huston, M.A. and A.C. McBride (2002), 'Evaluating the relative strengths of biotic versus abiotic controls on ecosystem processes', in M. Loreau, S. Naeem and P. Inchausti (eds), *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives*, Oxford, UK: Oxford University Press, pp. 47–60.

Illius, A.W. and T.G. O'Connor (1999), 'On the relevance of non-equilibrium concepts to arid and semi-arid grazing systems', *Ecological Applications*, **9**, 798–813.

Illius, A.W. and T.G. O'Connor (2000), 'Resource heterogeneity and ungulate population dynamics', *Oikos*, **89**, 283–294.

Inada, K.I. (1963), 'On a two-sector-model of economic growth: comments and a generalization', *Review of Economic Studies*, **30**, 119–127.

Ingham, R.E., J.A. Trofymow, E.R. Ingham and D.C. Coleman (1985), 'Interactions of bacteria, fungi, and their nematode grazers: effects on nutrient cycling and plant growth', *Ecological Monographs*, **55**, 119–140.

Intergovernmental Panel on Climate Change (IPCC) (2001), *Climate Change 2001. The Scientific Basis – Contribution of Working Group I to the Third Assessment Report of the IPCC*, Cambridge: Cambridge University Press.

IUPAC – International Union of Pure and Applied Chemistry (2003), www.iupac.org, searched on 6 October 2003.

Ives, A.R., K. Gross and J.L. Klug (1999), 'Stability and variability in competitive communities', *Science*, **286**, 542– 544.

Janssen, M.A., J.M. Anderies and B.H. Walker (2004), 'Robust strategies for managing rangelands with multiple stable attractors', *Journal of Environmental Economics and Management*, **47**, 140–162.

Johansson, P.-O (1987), *The Economic Theory and Measurement of Environmental Benefits*, Cambridge, UK: Cambridge University Press.

Johansson, P.-O (1999), 'Theory of economic valuation of environmental goods and services', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 747–757.

Johst K., M. Drechsler and F. Wätzold (2002), 'An ecological-economic modelling procedure to design effective and efficient compensation payments for the protection of species', *Ecological Economics*, **41**, 37–49.

Jöst, F., M. Quaas and J. Schiller (2004), 'Population growth and environmental deterioration: an intertemporal perspective', *Discussion Paper* No. 400, Department of Economics, University of Heidelberg.

Just, R.E. and R.D. Pope (1978), 'Stochastic specification of production functions and econometric implications', *Journal of Econometrics*, **7**(1), 67–86.

Just, R.E. and R.D. Pope (1979), 'Production function estimation and related risk considerations', *American Journal of Agricultural Economics*, **61**(2), 267–284.

Kant, I. (1996)[1785], 'Groundwork of the Metaphysics of Morals', in M.J. Gregor (transl. and ed.), *Practical Philosophy*, The Cambridge Edition of the Works of Immanuel Kant, Cambridge: Cambridge University Press, pp. 37–108.

Karp, L. and A. Pope III (1984), 'Range management under uncertainty', *American Journal of Agricultural Economics*, **66**, 437–446.

Kåberger, T. and B. Månsson (2001), 'Entropy and economic processes: physics perspectives', *Ecological Economics*, **36**, 165–179.

Keeler, E., M. Spence and R. Zeckhauser (1972), 'The optimal control of pollution', *Journal of Economic Theory*, **4**, 19–34.

Kinzig, A.P. and J. Harte (2000), 'Implications of endemics-area relationships for estimates of species extinctions', *Ecology*, **81**(12), 3305–3311.

Kinzig, A., S. Pacala and D. Tilman (2002), *The Functional Consequences of Biodiversity: Empirical Progress and Theoretical Extensions*, Princeton: Princeton University Press.

Kirchhoff, T. and L. Trepl (2001), 'Vom Wert der Biodiversität. Über konkurrierende politische Theorien in der Diskussion um Biodiversität', *Zeitschrift für angewandte Umweltforschung – Journal of Environmental Research*, special issue **13**/2001, 27–44.

Khalil, E.L. (1990), 'Entropy law and exhaustion of natural resources: Is Nicholas Georgescu-Roegen's paradigm defensible?', *Ecological Economics*, **2**, 163–178.

Klauer, B. (1999), 'Defining and achieving sustainable development', *International Journal of Sustainable Development and World Ecology*, **6**(2), 114–121.

Klauer, B. (2001), 'Welchen Beitrag können die Wirtschaftswissenschaften zum Erhalt der Biodiversität leisten?', *Zeitschrift für angewandte Umweltforschung – Journal of Environmental Research*, special issue **13**/2001, 59–70.

Kneese, A.V., R.U. Ayres and R.C. d'Arge (1972), *Economics and the Environment: A Materials Balance Approach*, Washington: Resources for the Future.

Kolstad, C.D. (2000), *Environmental Economics*, New York: Oxford University Press.

Kondepudi, D. and I. Prigogine (1998), *Modern Thermodynamics. From Heat Engines to Dissipative Structures*, New York: Wiley.

Koopmans, T.C. (1951a), 'Alternative proof of the substitution theorem for Leontief models in the case of three industries', in T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, New York: Wiley, pp. 147–154.

Koopmans, T.C. (1951b), 'Analysis of production as an efficient combination of activities', in T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, New York: Wiley, pp. 33–97.

Koopmans, T.C. (1957), 'Allocation of resources and the price system', in T.C. Koopmans (ed.), *Three Essays on the State of Economic Science*, New York: McGraw-Hill.

Kreps, D.M. (1990), *A Course in Microeconomic Theory*, New York: Harvester Wheatsheaf.

Krippendorff, K. (1986), *Information Theory. Structural Models for Qualitative Data*, Beverly Hills: Sage.

Krutilla, J.V. (1967), 'Conservation reconsidered', *Economic Review*, **57**, 777–786.

Krysiak, F.C. and D. Krysiak (2003), 'Production, consumption, and general equilibrium with physical constraints', *Journal of Environmental Economics and Management*, **46**, 513–538.

Kümmel, R. (1989), 'Energy as a factor of production and entropy as a pollutian indicator in macroeconomic modelling', *Ecological Economics*, **1**, 161–180.

Kümmel, R., D. Lindenberger and W. Eichhorn (2000), 'The productive power of energy and economic evolution', *Indian Journal of Applied Economics*, **8**, 231–262.

Kümmel, R. and U. Schüssler (1991), 'Heat equivalents of noxious substances: a pollution indicator for environmental accounting', *Ecological Economics*, **3**, 139–156.

Kümmel, R., W. Strassl, A. Grosser and W. Eichhorn (1985), 'Technical progress and energy dependent production functions', *Journal of Economics*, **45**, 285–311.

Kurz, H.D. and N. Salvadori (2003), 'Fund-flow versus flow-flow in production theory. Reflections on Georgescu-Roegen's contribution', *Journal of Economic Behavior and Organization*, **51**, 487–505.

Kurz, M. (1987), 'Myopic decision rules', in J. Eatwell, M. Milgate and P. Newman (eds), *The New Palgrave: A Dictionary of Economics*, Houndmills, UK and New York, NY, USA: Palgrave, pp. 578–581.

Lamprey, H.F. (1983), 'Pastoralism yesterday and today: The overgrazing problem', in F. Bourlière (ed.), *Tropical Savannas: Ecosystems of the World*, Amsterdam: Elsevier, pp. 643–666.

Landau, L.D. and E.E. Lifshitz (1980), *Statistical Physics*, 3rd edn., Oxford: Pergamon.

Lande, R. (1987), 'Extinction thresholds in demographic models of territorial populations', *American Naturalist*, **130**, 624–635.

Lande, R. (1993), 'Risk of population extinction from demographic and environmental stochasticity, and random catastrophs', *American Naturalist*, **142**, 911–927.

Lande, R., S. Engen, and B.-E. Saether (2003), *Stochastic Population Dynamics in Ecology and Conservation*, Oxford: Oxford University Press.

Landsberg, J. (1999), 'Response and effect: different reasons for classifying plant functional types under grazing', in D. Eldridge and D. Freudenberger (eds), *People and Rangelands: Building the Future*, Proceedings of the VI International Rangeland Congress, VI International Rangeland Congress, Townsville, Australia, pp. 911–915.

Lavorel, S. and E. Garnier (2002), 'Predicting the effects of environmental changes on plant community composition and ecosystem functioning: revisiting the Holy Grail', *Functional Ecology*, **16**, 545–556.

Lawton, J.H. (1994), 'What do species do in ecosystems?', *Oikos*, **71**, 367–374.

Lawton, J.H. and V.K. Brown (1993), 'Redundancy in ecosystems', in E.-D. Schulze and H.A. Mooney (eds), *Biodiversity and Ecosystem Functions*, Berlin, Heidelberg and New York: Springer, pp. 255–270.

Lehman, C.L. and D. Tilman (2000), 'Biodiversity, stability, and productivity in competitive communities', *The American Naturalist*, **156**, 534–552.

Leineweber, D.B., I. Bauer, H.G. Bock and J.P. Schlöder (2003), 'An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization. Part I: Theoretical aspects', *Computers & Chemical Engineering*, **27**, 157–166.

Lerch, A. (1995), 'Biologische Vielfalt – ein ganz normaler Rohstoff?', in J. Mayer (eds), *Eine Welt – eine Natur? Der Zugriff auf die biologische Vielfalt und die Schwierigkeiten global gerecht mit ihr umzugehen*, Loccum: Loccumer Protokolle Nr. 66/94, 33–62.

Lerch, A. (1996), *Verfügungsrechte und biologische Vielfalt. Eine Anwendung der ökonomischen Analyse der Eigentumsrechte auf die spezifischen Probleme genetischer Ressourcen*, Marburg: Metropolis.

Lerch, A. (1998), 'Property rights and biodiversity', *European Journal of Law and Economics*, **6**, 285–304.

Levin, S. (ed.) (2000), *Encyclopedia of Biodiversity*, New York: Academic Press.

Liiri, M., H. Setälä, J. Haimi, T. Pennanen and H. Fritze (2002), 'Relationship between soil microarthropod species diversity and plant growth does not change when the system is disturbed', *Oikos*, bf 96, 137–149.

Loreau, M. (1998a), 'Biodiversity and ecosystem functioning: a mechanistic model', *Proceedings of the National Academy of Sciences (USA)*, **95**, 5632–5636.

Loreau, M. (1998b), 'Separating sampling and other effects in biodiversity experiments', *Oikos*, **82**, 600–602.

Loreau, M. (2000), 'Biodiversity and ecosystem functioning: recent theoretical advances', *Oikos*, **91**, 3–17.

Loreau, M., A. Downing, M.C. Emmerson, A. Gonzalez, J. Hughes, P. Inchausti, J. Joshi, J. Norberg and O. Sala (2002a), 'A new look at the relationship between diversity and stability', in M. Loreau, S. Naeem and P. Inchausti (eds), *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives*, Oxford, UK: Oxford University Press, pp. 79–91.

Loreau, M. and A. Hector (2001), 'Partitioning selection and complementarity in biodiversity experiments', *Nature*, **412**, 72–76.

Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J.P. Grime, A. Hector, D.U. Hooper, M.A. Huston, D. Raffaelli, B. Schmid, D. Tilman and D.A. Wardle (2001), 'Biodiversity and ecosystem functioning: current knowledge and future challenges', *Science*, **294**, 804–808.

Loreau, M., S. Naeem, P. Inchausti (2002b), *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives*, Oxford: Oxford University Press.

Lozada, G.A. (1991), 'A defense of Nicholas Georgescu-Roegens's paradigm', *Ecological Economics*, **3**, 157– 160.

Lozada, G.A. (1995), 'Georgescu-Roegens's defence of classical thermodynamics revisited', *Ecological Economics*, **14**, 31–44.

Lundberg, P., E. Ranta and V. Kaitala (2000), 'Species loss leads to community closure', *Ecology Letters*, **3**(6), 465–468.

Luptacik, M. and U. Schubert (1982), *Optimal Economic Growth and the Environment. Economic Theory of Natural Resources*, Vienna: Physica.

MacArthur, R.H. (1955), ‘Fluctuations of animal populations and a measure of community stability’, *Ecology*, **36**, 533–536.

MacArthur, R.H. and E.O. Wilson (1967), *The Theory of Island Biogeography*, Princeton: Princeton University Press.

MacMahon, J.A. (1997), ‘Ecological restoration’, in G.K. Meffe and C.R. Carroll (eds), *Principles of Conservation Biology*, 2nd edn., Sunderland, MA: Sinauer Associates, pp. 479–511.

Mageau, M.T., R. Costanza and R.E. Ulanowicz (1995), ‘The development and initial testing of a quantitative assessment of ecosystem health’, *Ecosystem Health*, **1**, 201–213.

Magurran, A.E. (2004), *Measuring Biological Diversity*, Malden, MA, USA and Oxford, UK: Blackwell.

Mäler, K.-G. (1974), *Environmental Economics: A Theoretical Inquiry*, Baltimore, MD, USA and London, UK: Johns Hopkins University Press.

Malthus, T.R. (1798), *An Essay on the Principle of Population*, London: printed for J. Johnson, in St. Paul’s church-yard.

Mann, S. (1998), *Nachwachsende Rohstoffe*, Stuttgart: Ulmer.

Marggraf, R. and S. Streb (1997), *Ökonomische Bewertung der natürlichen Umwelt. Theorie, politische Bedeutung, ethische Diskussion*, Heidelberg and Berlin: Spektrum.

Margules, C.R., A.O. Nicholls and R.L. Pressey (1988), ‘Selecting networks of reserves to maximise biological diversity’, *Biological Conservation*, **43**, 63–76.

Martinez-Alier, J. (1997), ‘Some issues in agrarian and ecological economics: in memory of Georgescu-Roegen’, *Ecological Economics*, **22**, 225–238.

Mas-Colell, A., M.D. Whinston and J.R. Green (1995), *Microeconomic Theory*, New York: Oxford University Press.

Mateo, N., W.F. Nader and G. Tamayo (2000), ‘Bioprospecting’, in Levin (ed.), *Encyclopedia of Biodiversity*, New York: Academic Press.

Maurer, U. (1995), *Nutzungspotential und Landdegradation im semi-ariden Namibia: ein Beitrag zur Desertifikationsforschung*, PhD-Thesis, University of Göttingen.

May, R.M. (1972), ‘Will a large complex system be stable?’, *Nature*, **238**, 413–414.

May, R.M. (1974), *Stability and Complexity in Model Ecosystems*, 2nd edn., Princeton: Princeton University Press.

May, R.M. (1975), ‘Patterns of species abundance and diversity’, in M.L. Cody and J.M. Diamond (eds), *Ecology and Evolution of Communities*, Cambridge, MA: Harvard University Press, pp. 81–120.

May, R.M. (1990), ‘Taxonomy as destiny’, *Nature*, **347**, 129–130.

May, R.M., J.H. Lawton and N.E. Stork (1995), ‘Assessing extinction rates’, in R.M. May and J.H. Lawton (eds), *Extinction rates*, Oxford, UK: Oxford University Press, pp. 1–24.

Mayers, D. and C.W. Smith Jr. (1990), ‘On the corporate demand for insurance: evidence from the reinsurance market’, *Journal of Business*, **63**(1), 19–40.

Mayumi, K. (2001), *The Origins of Ecological Economics. The Bioeconomics of Georgescu-Roegen*, London: Routledge.

Mayumi, K. and J.M. Gowdy (eds) (1999), *Bioeconomics and Sustainability. Essays in Honor of Nicholas Georgescu-Roegen*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

McArthur, D. and J.L. Dillon (1971), 'Risk, utility and stocking rate', *Australian Journal of Agricultural Economics*, **15**(1), 20–35.

McCann, K.S. (2000), 'The diversity-stability debate', *Nature*, **405**, 228–233.

McCann, K.S., A. Hastings and G.R. Huxel (1998), 'Weak trophic interactions and the balance of nature', *Nature*, **395**, 794–798.

McGrady-Steed, J., P.M. Harris and P.J. Morin (1997), 'Biodiversity regulates ecosystem predictability', *Nature*, **390**, 162–164.

McIntosh, R.P. (1967), 'An index of diversity and the relation of certain concepts to diversity', *Ecology*, **48**, 392–404.

McNaughton, S.J. (1977), 'Diversity and stability of ecological communities: a comment on the role of empiricism in ecology', *American Naturalist*, **111**, 515–525.

McNaughton, S.J. (1985), 'Ecology of a grazing ecosystem: the Serengeti', it Ecological Monographs, **55**, 259–294.

McNaughton, S.J. (1993), 'Biodiversity and function of grazing ecosystems', in E.-D. Schulze and H.A. Mooney (eds), *Biodiversity and Ecosystem Function*, Berlin, Heidelberg and New York: Springer, pp. 361–383.

McNeely, J.A. (1988), *Economics and Biological Diversity: Developping and Using Economic Incentives to Conserve Biological Resources*, commissioned by the International Union for Conservation of Nature and Natural Resources (IUCN), Gland.

McNeely, J.A. et al. (1990), *Conserving the World's Biological Diversity*, commissioned by the International Union for Conservation of Nature and Natural Resources, World Resources Institute, Conservation International, World Wildlife Fund-US, and the World Bank, Washington DC: World Bank.

McKenzie, L.W. (1959), 'On the existence of general equilibrium for a competitive market', *Econometrica*, **27**, 54–71.

McKenzie, L.W. (1961), 'On the existence of general equilibrium for a competitive market: some corrections', *Econometrica*, **29**, 247–248.

Meadows, D.H., D.L. Meadows, J. Randers and W.W. Behrens (1972), *The Limits to Growth*, New York: Universe.

Mendelsohn, J., A. Jarvis, C. Roberts and T. Robertson (2002), *Atlas of Namibia: A Portrait of the Land and it's People*, Cape Town: David Philip.

Metrick, A. and M.L. Weitzman (1996), 'Patterns of behavior in endangered species preservation', *Land Economics*, **72**(1), 1–16.

Metrick, A. and M.L. Weitzman (1998), 'Conflicts and choices in biodiversity preservation', *Journal of Economic Perspectives*, **12**(3), 21–34.

Michaelis, P. (1992), 'Global warming: efficient policies in the case of multiple pollutants', *Environmental and Resource Economics*, **2**, 61–77.

Michaelis, P. (1999), 'Sustainable greenhouse policies: the role of non-CO_2 gases', *Structural Change and Economic Dynamics*, **10**, 239–260.

Millennium Ecosystem Assessment (2005), *Ecosystems and Human Well-Being: Synthesis Report*, Washington DC: Island Press.

Mirowski, P. (1989), *More Heat than Light. Economics as Social Physics, Physics as Nature's Economics*, Cambridge, UK and New York, NY, USA: Cambridge University Press.

Mooney, H.A. and P.R. Ehrlich (1997), 'Ecosystem services: a fragmentary history', in G.C. Daily (ed.), *Nature's Services. Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press, pp. 11–19.

Moran, D. and D.W. Pearce (1997), 'The economics of biodiversity', in H. Folmer and T. Tietenberg (eds), *The International Yearbook of Environmental and Resource Economics 1997/1998*, Cheltenham, UK and Lyme, USA: Edward Elgar, pp. 82–113.

Moslener, U. and T. Requate (2001), 'Optimal abatement strategies for various interacting greenhouse gases. Is the global warming potential a useful indicator?', *Discussion Paper* No. 360, Department of Economics, University of Heidelberg.

Mouquet, N., J.L. Moore and M. Loreau (2002), 'Plant species richness and community productivity: why the mechanism that promotes coexistence matters', *Ecology Letters*, **5**, 56–65.

Munasinghe, M. (1992), 'Biodiversity protection policy: Environmental valuation and distribution issues', *Ambio*, **21**, 227–236.

Muradian, R. (2001), 'Ecological thresholds: a survey', *Ecological Economics*, **38**, 7–24.

Müller, B., K. Frank and C. Wissel (forthcoming), 'Relevance of rest periods in non-equilibrium rangeland systems. A modelling analysis', *Agricultural Systems*.

Myers, N. (1979), *The Sinking Ark*, Oxford: Pergamon.

Myers, N. (1983), *A Wealth of Wild Species: Storehouse for Human Welfare*, Boulder: Westview.

Myers, N. (1989), 'Loss of biological diversity and its potential impact on agriculture and food production', in D. Pimentel and C.W. Hall (eds), *Food and Natural Resources*, San Diego: Academic Press, pp. 49–68.

Myers, N. (1995), 'Tropical deforestation: population, poverty and biodiversity', in T.M. Swanson (ed.), *The economics and ecology of biodiversity decline. The forces driving global change*, Cambridge: Cambridge University Press, pp. 111–122.

Myers, N. (1997), 'Biodiversity's genetic library', in G.C. Daily (ed.), *Nature's Services: Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press, pp. 255–273.

Nabhan, G.P. and S.L. Buchmann (1997), 'Services provided by pollinators', in G.C. Daily (ed.), *Nature's Services: Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press, pp. 133–150.

Nader, W.F. and B. Hill (1999), *Der Schatz im Tropenwald. Biodiversität als Inspirations- und Innovationsquelle*, Aachen: Shaker.

Naeem, S. (1998), ‘Species redundancy and ecosystem reliability’, *Conservation Biology*, **12**, 39–45.

Naeem, S. (2002), ‘Ecosystem consequences of biodiversity loss: the evolution of a paradigm’, *Ecology*, **83**, 1537–1552.

Naeem, S. and S. Li (1997), ‘Biodiversity enhances reliability’, *Nature*, **390**, 507–509.

Naeem, S., L.J. Thompson, S.P. Lawler, J.H. Lawton and R.M. Woodfin (1995), ‘Empirical evidence that declining species diversity may alter the performance of terrestrial ecosystems’, *Philosophical Transactions of the Royal Society of London*, **B 347**, 249–262.

Nakićenović, N., P.V. Gilli and R. Kurz (1996), ‘Regional and global exergy and energy efficiencies’, *Energy*, **21**, 223–237.

Naylor, R.L. and P.R. Ehrlich (1997), ‘Natural pest control services and agriculture’, in G.C. Daily (ed.), *Nature’s Services: Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press, pp. 151–174.

Nehring, K. and C. Puppe (2002), ‘A theory of diversity’, *Econometrica*, **70**, 1155–1198.

Nehring, K. and C. Puppe (2004), ‘Modelling phylogenetic diversity’, *Resource and Energy Economics*, **26**, 205–235.

Neumayer, E. (2003), *Weak Versus Strong Sustainability: Exploring the Limits of Two Opposing Paradigms*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Niehans, J. (1990), *A History of Economic Theory. Classic Contributions 1720–1980*, Baltimore, MD, USA and London, UK: Johns Hopkins University Press.

Niklaus, P.A., E. Kandeler, P.W. Leadley, B. Schmid, D. Tscherko and C. Körner (2001), ‘A link between plant diversity, elevated CO_2 and soil nitrate’, *Oecologia*, **127**, 540–548.

Noll, R. and J. Trijonis (1971), ‘Mass balance, general equilibrium, and environmental externalities’, *American Economic Review*, **61**, 730–735.

Norgaard, R.B. (1981), ‘Sociosystem and ecosystem coevolution in the Amazon’, *Journal of Environmental Economics and Management*, **8**, 238–254.

Norgaard, R.B. (1984), ‘Coevolutionary development potential’, *Land Economics*, **60**(2), 160–173.

Norgaard, R.B. (1985), ‘Environmental economics: an evolutionary critique and a plea for pluralism’, *Journal of Environmental Economics and Management*, **12**, 382–394.

Norgaard, R.B. (1986), ‘Thermodynamic and economic concepts as related to resource-use policies: synthesis’, *Land Economics*, **62**, 325–328.

Norgaard, R.B. (1994), ‘The coevolution of environmental and economic systems and the emergence of unsustainablility’ in R.W. England (ed.), *Evolutionary Concepts in Contemporary Economics*, Ann Arbor: University of Michigan Press, pp. 213–225.

Noy-Meir, I. (1982), ‘Stability of plant-herbivore models and possible applications to savanna’, in B. Huntley and B. Walker (eds), *Ecology and Tropical Savannas*, Berlin, Heidelberg and New York: Springer.

Nunes, P.A.L.D. and J.C.J.M. van den Bergh (2001), 'Economic valuation of biodiversity: sense or nonsense?', *Ecological Economics*, **39**, 203–222.

Nunes, P.A.L.D., J.C.J.M. van den Bergh and P. Nijkamp (2003), *The Ecological Economics of Biodiversity. Methods and Policy Applications*, Cheltenham, UK, and Northampton, MA, USA: Edward Elgar.

O'Connor, T.G. and T. Everson (1998), 'Population dynamics of perennial grasses in African savanna and grasslands', in G. Cheplik (ed.), *Population Biology of Grasses*, Cambridge: Cambridge University Press, pp. 333–365.

Odum, E. (1953), *Fundamentals of Ecology*, Philadelphia: Saunders.

Odum, H.T. (1971), *Environment, Power and Society*, New York: Wiley.

Oldfield, S. (1992), 'Plant use', in B. Groombridge (ed.), *Global Biodiversity: Status of the World's Living Resources. A Report Compiled by the World Conservation Monitoring Centre*, London: Chapman & Hall, pp. 331–358.

Ozdogan, S. and M. Arikol (1981), 'Energy and exergy analysis of selected Turkish industries', *Energy*, **18**, 73–80.

Pacala, S.W. and D.H. Deutschman (1995), 'Details that matter: the spatial distribution of individual trees maintains forest ecosystem function', *Oikos*, **74**, 357–365.

Paine, R.T. (1992), 'Food-web analysis through field measurements of per capita interaction strengths', *Nature*, **355**, 73–75.

Patterson, M.G. (1998), 'Commensuration and theories of value in ecological economics', *Ecological Economics*, **25**, 105–125.

Pearce, D.W. (1993), *Economic Values and the Natural World*, London: Earthscan.

Pearce, D.W., E. Barbier and A. Markandya (1990), *Sustainable Development: Economics and Environment in the Third World*, Aldershot, UK, and Brookfield, USA: Edward Elgar.

Pearce, D.W. and D. Moran (1994), *The Economic Value of Biodiversity*, London: Earthscan.

Pearce, D.W. and R.K. Turner (1990), *Economics of Natural Resources and the Natural Environment*, New York: Harvester Wheatsheaf.

Peet, R.K. (1974), 'The measurement of species diversity', *Annual Review of Ecological Systems*, **5**, 285–307.

Perrings, C. (1987), *Economy and Environment: A Theoretical Essay on the Interdependence of Economic and Environmental Systems*, Cambridge: Cambridge University Press.

Perrings, C. (1995a), 'Biodiversity conservation as insurance', in T.M. Swanson (ed.), *The economics and ecology of biodiversity decline. The forces driving global change*, Cambridge: Cambridge University Press, pp. 69–77.

Perrings, C. (1995b), 'Ecological resilience in the sustainability of economic development', *Economie Appliquée*, **48**, 121–142.

Perrings, C. (1995c), 'Preface', in C. Perrings, K.-G. Mäler, C. Folke, C.S. Holling and B.-O. Jansson (eds), *Biodiversity Loss. Economic and Ecological Issues*, Cambridge: Cambridge University Press, pp. xi-xii.

Perrings, C. (1997), 'Stress, shock and the sustainability of resource use in semi-arid environments', in C. Perrings, *Economics of Ecological Resources*, Cheltenham, UK and Lyme, USA: Edward Elgar, pp. 142–175.

Perrings, C. (2001), 'Resilience and sustainability', in H. Folmer, H.L. Gabel, S. Gerking and A. Rose (eds), *Frontiers of Environmental Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 319–341.

Perrings, C., K.-G. Mäler, C. Folke, C.S. Holling and B.-O. Jansson (eds) (1995a), *Biodiversity Loss. Economic and Ecological Issues*, Cambridge: Cambridge University Press.

Perrings, C., K.-G. Mäler, C. Folke, C.S. Holling and B.-O. Jansson (1995b), 'Introduction: Framing the problem of biodiversity loss', in C. Perrings, K.-G. Mäler, C. Folke, C.S. Holling and B.-O. Jansson (eds), *Biodiversity Loss. Economic and Ecological Issues*, Cambridge: Cambridge University Press, pp. 1–17.

Perrings, C. and B.H. Walker (1997), 'Biodiversity, resilience and the control of ecological-economic systems: the case of fire-driven rangelands', *Ecological Economics*, **22**, 73–83.

Perrings, C. and B.H. Walker (2004), 'Conservation in the optimal use of rangelands', *Ecological Economics*, **49**, 119–128.

Petchey, O.L., P.T. McPhearson, T.M. Casey and P.J. Morin (1999), 'Environmental warming alters food-web structure and ecosystem function', *Nature*, **402**, 69–72.

Peterson, G., C.R. Allen and C.S. Holling (1998), 'Ecological resilience, biodiversity and scale', *Ecosystems*, **1**, 6–18.

Pethig, R. (1979), 'Environmental management in general equilibrium. A new incentive compatible approach', *International Economic Review*, **20**, 1–27.

Pethig, R. (2003), 'The "materials balance approach" to pollution: its origin, implications and acceptance', *Discussion Paper 105–03*, University of Siegen.

Pethig, R. (2006), 'Non-linear production, abatement, pollution and materials balance reconsidered', *Journal of Environmental Economics and Management*, **51**(2), 185–204.

Pezzey, J. (1992), *Sustainable Development Concepts. An Economic Analysis*, Washington DC: World Bank.

Pfisterer, A.B., J. Joshi, B. Schmid and M. Fischer (2004), 'Rapid decay of diversity–productivity relationships after invasion in experimental plant communities', *Basic and Applied Ecology*, **5**, 5–14.

Pfisterer, A.B. and B. Schmid 2002, 'Diversity-dependent production can decrease the stability of ecosystem functioning', *Nature*, **416**, 84–86.

Pielou, E.C. (1975), *Ecological Diversity*, New York: Wiley.

Pigou, A.C. (1912), *Wealth and Welfare*, London: Macmillan.

Pigou, A.C. (1920), *The Economics of Welfare*, London: Macmillan.

Pimm, S.L. (1984), 'The complexity and stability of ecosystems', *Nature*, **307**, 321—326.

Pindyck, R.S. (1991), 'Irreversibility, uncertainty, and investment', *Journal of Economic Literature*, **29**(3), 1110–1152.

Pirscher, F. (1997), *Möglichkeiten und Grenzen der Integration von Artenvielfalt in die ökonomische Bewertung vor dem Hintergrund ethischer Normen*, Frankfurt: Lang.

Plourde, C.G. (1972), 'A model of waste accumulation and disposal', *Canadian Journal of Economics*, **5**, 119–125.

Plotkin, M.J. (1988), 'The outlook for new agricultural and industrial products from the tropics', in E.O. Wilson (ed.), *BioDiversity*, Washington, DC: National Academy Press, pp. 106–116.

Polasky, S., J.D. Camm and B. Garber-Yonts (2001), 'Selecting biological reserves cost-effectively: an application to terrestrial vertebrate conservation in Oregon', *Land Economics*, **77**, 68–78.

Polasky, S. and A. Solow (1995), 'On the value of a collection of species', *Journal of Environmental Economics and Management*, **29**, 298–181.

Polasky, S., A. Solow and J. Broadus (1993), 'Searching for uncertain benefits and the conservation of biological diversity', *Environmental and Resource Economics*, **3**, 171–181.

Pommerehne, W.W. (1987), *Präferenzen für öffentliche Güter: Ansätze zu ihrer Erfassung*, Tübingen: Mohr-Siebeck.

Powell, J., K. Turner, M. Peters, and B. Strobl (2002), 'Economics of waste', in K. Bisson and J. Proops (eds), *Waste in Ecological Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 73–100.

Prigogine, I. (1962), *Introduction to Non-Equilibrium Thermodynamics*, New York: Wiley.

Prigogine, I. (1967), *Thermodynamics of Irreversible Processes*, New York: Interscience.

Proops, J. (1985), 'Thermodynamics and economics: from analogy to physical functioning', in van W. van Gool and J.J.C. Bruggink (eds), *Energy and Time in the Economic and Physical Sciences*, Amsterdam: North-Holland.

Proops, J. (1987), 'Entropy, information and confusion in the social sciences', *Journal of Interdisciplinary Economics*, **1**, 224–242.

Purvis, A. and A. Hector (2000), 'Getting the measure of biodiversity', *Nature*, **405**, 212–219.

Pratt, J.W. (1964), 'Risk aversion in the small and in the large', *Econometrica*, **32**(1–2), 122–136.

Quaas M. and S. Baumgärtner (2006), 'Natural vs. financial insurance in the management of public-good ecosystems', *Working Paper*, Department of Economics, University of Heidelberg.

Quaas M., S. Baumgärtner, C. Becker, B. Müller and K. Frank (2004), 'Uncertainty and sustainability in the management of semi-arid rangelands', *Discussion Paper* No. 414, Department of Economics, University of Heidelberg.

Quirk, M. (2002), 'Managing grazing', in A.C. Grice and K.C. Hodgkinson (eds), *Global Rangelands: Progress and Prospects*, Oxford: Oxford University Press, pp. 131–145.

Randall, A. (1988), 'What mainstream economists have to say about the value of biodiversity', in E.O. Wilson (ed.), *BioDiversity*, Washington, DC: National Academy Press, pp. 217–223.

Rao, C.R. (1982), 'Diversity and dissimilarity coefficients: a unified approach', *Theoretical Population Ecology*, **21**, 24–43.

Rapport, D.J., R. Costanza and A. McMichael (1998), 'Assessing ecosystem health: challenges at the interface of social, natural, and health sciences', *Trends in Ecology and Evolution*, **13**, 397–402.

Rausser, G.C. and A.A. Small (2000), 'Valuing research leads: bioprospecting and the conservation of genetic resources', *Journal of Political Economy*, **108**, 173–206.

Reich, P.B. et al. (2001), 'Plant diversity enhances ecosystem responses to elevated CO_2 and nitrogen deposition', *Nature*, **410**, 809–812.

Rényi, A. (1961), 'On measures of entropy and information', in J. Neyman (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, Berkeley: University of California Press, pp. 547–561.

Ricklefs, R.E. and G.L. Miller (2000), *Ecology*, 4th edn., New York: Freeman.

Robbins, L. (1932), *An Essay on the Nature and Significance of Economic Science*, London: Macmillan.

Rodriguez, A. and R.G. Taylor (1988), 'Stochastic modeling of short-term cattle operations', *American Journal of Agricultural Economics*, **70**, 121–132.

Rosenzweig, M.L. (1995), *Species Diversity in Space and Time*, Cambridge, UK: Cambridge University Press.

Røpke, I. (2004), 'The early history of modern ecological economics', *Ecological Economics*, **50**, 293–314.

Ruth, M. (1993), *Integrating Economics, Ecology and Thermodynamics*, Dordrecht: Kluwer.

Ruth, M. (1995a), 'Technology change in the US iron and steel production', *Resources Policy*, **21**, 199–214.

Ruth, M. (1995b), 'Thermodynamic constraints on optimal depletion of copper and aluminum in the United States: A dynamic model of substitution and technical change', *Ecological Economics*, **15**, 197–213.

Ruth, M. (1995c), 'Thermodynamic implications for natural resource extraction and technical change in US copper mining', *Environmental and Resource Economics*, **6**, 187–206.

Ruth, M. (1999), 'Physical principles in environmental economic analysis', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 855–866.

Sala, O.E., W.K. Lauenroth, S.J. McNaughton, G. Rusch and X. Zhang (1996), 'Biodiversity and ecosystem function in grasslands', in H.A. Mooney, J.H. Cushman, E. Medina, O.E. Sala and E.-D. Schulze (eds), *Functional Roles of Biodiversity: A Global Perspective*, Chichester, UK: John Wiley, pp. 129–149.

Saltelli, A., K. Chan and E.M. Scott (eds) (2000), *Sensitivity Analysis*, New York: Wiley.

Samuelson, P.A. (1947), *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

Samuelson, P.A. (1951), 'Abstract of a theorem concerning substitutability in open Leontief models', in T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, New York: Wiley, pp. 142–146.

Samuelson, P.A. (1960a), 'Structure of minimum equilibrium systems', in R.W. Pfouts (ed.), *Essays in Economics and Econometrics: A Volume in Honor of Harold Hotelling*, Chapel Hill, NC: University of North Caroline Press, pp. 1–33.

Samuelson, P.A. (1960b), 'An extension of the Le Chatelier Principle', *Econometrica*, **28**, 368–379.

Sandelin, B. (1976), 'On the origin of the Cobb Douglas production function', *Economy and History*, **19**, 117–123.

Sandford, S. (1982), 'Pastoral strategies and desertification: opportunism and conservation in dry lands', in B. Spooner and H. Mann (eds), *Desertification and Development*, London: Academic Press, pp. 61–80.

Sandford, S. (1994), 'Improving the efficiency of opportunism: New directions for pastoral development', in I. Scoones (ed.), *Living with Uncertainty. New Directions in Pastoral Development in Africa*, London: Intermediate Technology Publications, pp. 174–182.

Schaeffer, R. and R.M. Wirtshafter (1992), 'An exergy analysis of the Brazilian economy: from energy products to final use', *Energy*, **17**, 841–855.

Schefold, B. (2001), 'Ökonomische Bewertung der Natur aus dogmengeschichtlicher Perspektive. Eine Skizze', in F. Beckenbach et al. (eds), *Jahrbuch Ökologische Ökonomik, Band 2: Ökonomische Naturbewertung*, Marburg: Metropolis, pp. 17–61.

Schindler, D.W., M.A. Turner, M.P. Stainton and G.A. Linsey (1986), 'Natural sources of acid neutralizing capacity in low alkalinity lakes of the Precambrian shield', *Science*, **232**, 844–847.

Schläpfer, B. Pfisterer and B. Schmid (2005), 'Non-random species extinction and plant production: implications for ecosystem functioning', *Journal of Applied Ecology*, **42**, 13–24.

Schläpfer, F. M. Tucker, I. Seidl (2002), 'Returns from hay cultivation in fertilized low diversity and non-fertilized high diversity grassland', *Environmental and Resource Economics*, **21**, 89–100.

Schmid, B., A. Hector, M.A. Huston, P. Inchausti, I. Nijs, P.W. Leadley and D. Tilman (2002) 'The design and analysis of biodiversity experiments', in M. Loreau, S. Naeem and P. Inchausti (eds), *Biodiversity and Ecosystem Functioning. Synthesis and Perspectives*, Oxford, UK: Oxford University Press, pp. 61–75.

Schulze, E.-D. and H.A. Mooney (eds) (1993), *Biodiversity and Ecosystem Function*, Berlin, Heidelberg and New York: Springer.

Schulze, W.D. (1974), 'The optimal use of non-renewable resources: The theory of extraction', *Journal of Environmental Economics and Management*, **1**, 53–73.

Schumpeter, J.A. (1954), *History of Economic Analysis*, London: Oxford University Press.

Scoones, I. (1994), 'New directions for pastoral development in Africa', in I. Scoones (ed.), *Living with Uncertainty. New Directions in Pastoral Development in Africa*, London: Intermediate Technology Publications, pp. 1–36.

Sedjo, R.A. and R.D. Simpson (1995), 'Property rights, externalities and biodiversity', in T.M. Swanson (ed.), *The economics and ecology of biodiversity decline. The forces driving global change*, Cambridge: Cambridge University Press, pp. 79–88.

Seidl, I. and J. Gowdy (1999), 'Monetäre Bewertung von Biodiversität: Grundannahmen, Schritte, Probleme und Folgerungen', *GAIA*, **8**, 102–112.

Shannon, C.E. (1948), 'A mathematical theory of communication', *Bell System Technical Journal*, **27**, 379–423 and 623–656.

Shannon, C.E. and W. Weaver (1949), *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press.

Shephard, R.W. (1970), *Theory of Cost and Production Functions*, Princeton: Princeton University Press.

Shogren, J.F. and T.D. Crocker (1999), 'Risk and its consequences', *Journal of Environmental Economics and Management*, **37**, 44–51.

Shogren, J.F., J. Tschirhart, T. Anderson, A.W. Ando, S.R. Beissinger, D. Brookshire, G.M. Brown Jr., D. Coursey, R. Innes, S.M. Meyer and S. Polasky (1999), 'Why economics matters for endangered species protection', *Conservation Biology*, **13**, 1257–1261.

Shouse, B. (2002), 'Endangered Species Act. Cherished concepts faltering in the field', *Science*, **296**, 1219–1220.

Siebert, H. (2004), *Economics of the Environment. Theory and Policy*, 6th edn., Berlin, Heidelberg and New York: Springer.

Simberloff, D. (1998), 'Flagships, umbrellas, and keystones: is single-species management passé in the landscape era?', *Biological Conservation*, **83**, 247–257.

Simpson, E.H. (1949), 'Measurement of diversity', *Nature*, **163**, 688.

Simpson, R.D., R.A. Sedjo and J.W. Reid (1996), 'Valuing biodiversity for use in pharmaceutical research', *Journal of Political Economy*, **104**, 163–185.

Smale, M., J. Hartell, P.W. Heisey and B. Senauer (1998), 'The contribution of genetic resources and diversity to wheat production in the Punjab of Pakistan', *American Journal of Agricultural Economics*, **80**, 482–493.

Smith, E. and D. Foley (2004), 'Classical thermodynamics and economic general equilibrium theory', *New School University, Center for Economic Policy Analysis Working Paper.*

Smith, F.D.M., G.C. Daily and P.R. Ehrlich (1995), 'Human population dynamics and biodiversity loss', in T.M. Swanson (ed.), *The economics and ecology of biodiversity decline. The forces driving global change*, Cambridge: Cambridge University Press, pp. 125–114.

Smith, K.V. (1996), *Estimating Economic Values for Nature: Methods for Non-Market Valuation*, Cheltenham, UK and Brookfield, USA: Edward Elgar.

Smith, V.L. (1977), 'Control theory applied to natural and environmental resources. An exposition', *Journal of Environmental Economics and Management*, **4**, 1–24.

Smulders, S. (1999), 'Endogenous growth theory and the environment', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 610–621.

Söllner, F. (1997), 'A reexamination of the role of thermodynamics for environmental economics', *Ecological Economics*, **22**, 175–201.

Solow, R.M. (1956), 'A contribution to the theory of economic growth', *Quarterly Journal of Eocnomics*, **70**, 65–94.

Solow, R.M. (1974), 'Intergenerational equity and exhaustible resources', *Review of Economic Studies*, **41** *(Symposium on the Economics of Exhaustible Resources)*, 29–45.

Solow, R.M. (1997), 'Georgescu-Roegen versus Solow/Stiglitz. Reply', *Ecological Economcis*, **22**, 267–268.

Solow, A., S. Polasky and J. Broadus (1993), 'On the measurement of biological diversity', *Journal of Environmental Economics and Management*, **24**, 60–68.

Soulé, M.E. (ed.) (1986), *Conservation Biology: The Science of Scarcity and Diversity*, Sunderland, MA: Sinauer.

Soulé, M.E. and B.A. Wilcox (eds) (1980), *Conservation Biology: An Evolutionary-Ecological Perspective*, Sunderland, MA: Sinauer.

Sousa, T. and T. Domingo (forthcoming), 'Is neoclassical economics formally valid? An approach based on an analogy between thermodynamics and economics', *Ecological Economics*.

Spash, C. (1999), 'The development of environmental thinking in economics', *Environmental Values*, **8**, 413–435.

Spreng, D. (1993), 'Possibilities for substitution between energy, time and information', *Energy Policy*, **21**, 13–23.

Statistisches Bundesamt (1997), *Beiträge zu den Umweltökonomischen Gesamtrechnungen, Band 1: Physische Input-Output-Tabellen 1990*, Wiesbaden: Statistisches Bundesamt.

Stephan, T., F. Jeltsch, T. Wiegand and H.A. Breiting (1998), 'Sustainable farming at the edge of the Namib: analysis of land use strategies by a computer simulation model' in J.L. Uso, C.A. Brebbia and H. Power (eds), *Ecosystems and Sustainable Development*, Southampton: Computational Mechanics, pp. 41–50.

Stephan, T., F. Jeltsch, T. Wiegand, C. Wissel and H.A. Breiting (1996), 'Die Gamis-Farm in Namibia: Analyse einer Weidewirtschaftsstrategie mit Hilfe eines Computersimulationsmodells', *Verhandlungen der Gesellschaft für Ökologie*, **26**, 455–462.

Stephan, T., F. Jeltsch, T. Wiegand, C. Wissel and H.A. Breiting (1998), 'A simulation model analyzing land use strategies on Gamis Farm, Namibia', in *Proceedings of the 15th International AFSRE-Symposium, Pretoria, SA*, pp. 306–314.

Stiglitz, J. (1974), 'Growth with exhaustible natural resources: efficient and optimal growth paths', *Review of Economic Studies*, **41** (Symposium on the Economics of Exhaustible Resources), 123–152.

Stiglitz, J.E. (1997), 'Georgescu-Roegen versus Solow/Stiglitz. Reply', *Ecological Economcis*, **22**, 269–270.

Strasburger, E.F. Noll, H. Schenck and A.F.W. Schimper (1991), *Lehrbuch der Botanik für Hochschulen*, 33. Auflage neubearbeitet von P. Sitte, H. Ziegler, F. Ehrendorfer and A. Bresinsky, Stuttgart: Fischer.

Stuth, J. and G.E. Maraschin (2000), 'Sustainable management of pasture and rangelands', in: G. Lemaire, J. Hodgson, A. Moraes, P.C.F. Carvalho and C. Nabinger (eds), *Grassland Ecophysiology and Grazing Ecology*, Wallingford, UK: CAB International, pp. 339–354.

Sullivan, S. and R. Rhode (2002), 'On non-equilibrium in arid and semi-arid grazing systems', *Journal of Biogeography*, **29**, 1595–1618.

Swan, T.W. (1956), 'Economic growth and capital accumulation', *Economic Record*, **32**, 334–361.

Swanson, T.M. (1994), *The International Regulation of Extinction*, London: Macmillan.

Swanson, T.M. (ed.) (1995a), *The economics and ecology of biodiversity decline. The forces driving global change*, Cambridge: Cambridge University Press.

Swanson, T.M. (1995b), 'Why does biodiversity decline? The analysis of forces for global change', in T.M. Swanson (ed.), *The economics and ecology of biodiversity decline. The forces driving global change*, Cambridge: Cambridge University Press, pp. 1–9.

Swanson, T.M. (1996), 'The reliance of northern economies on southern biodiversity: biodiversity as information', *Ecological Economics*, **17**, 1–8.

Swanson, T.M. and T. Goeschl (2000a), 'Genetic use restriction technologies and the diffusion of yield gains to developing countries', *Journal of International Development*, **12**(8), 1159–1178.

Swanson, T.M. and T. Goeschl (2000b), 'Property rights issues involving plant genetic resources: implications of ownership for economic efficiency', *Ecological Economics*, **32**, 75–92.

Swanson, T.M. and T. Goeschl (2003), 'Pests, plagues, and patents', *Journal of the European Economic Association*, **1**(2–3), 561–575.

Swanson, T.M. et al. (1992), 'Biodiversity and Economics', in B. Groombridge (ed.), *Global Biodiversity: Status of the World's Living Resources. A Report Compiled by the World Conservation Monitoring Centre*, London: Chapman & Hall, pp. 407–438.

Symstad, A.J. and D. Tilman (2001), 'Diversity loss, recruitment limitation, and ecoystem functioning: lessons learned from a removal experiment', *Oikos*, **92**, 424–435.

Szargut, J., D.R. Morris and F.R. Steward (1988), *Exergy Analysis of Thermal, Chemical, and Metallurgical Processes*, New York: Hemisphere.

Takayama, A. (1985), *Mathematical Economics*, 2nd edn., Cambridge: Cambridge University Press.

ten Kate, K. (1995), *Biopiracy or Green Petroleum? Expectations and Best Practice in Bioprospecting*, London: Overseas Development Administration.

Tietenberg, T. (2003), *Environmental and Natural Resource Economics*, 6th edn., Addison-Wesley.

Tilman, D. (1996), 'Biodiversity: population versus ecosystem stability', *Ecology*, **77**, 350–363.

Tilman, D. (1997a), 'Biodiversity and ecosystem functioning', in G.C. Daily (ed.), *Nature's Services. Societal Dependence on Natural Ecosystems*, Washington, DC: Island Press, pp. 93–112.

Tilman, D. (1997b), 'Distinguishing between the effects of species diversity and species composition', *Oikos*, **80**, 185.

Tilman, D. (1999), 'The ecological consequences of changes in biodiversity: a search for general principles', *Ecology*, **80**, 1455–1474.

Tilman, D. and J.A. Downing (1994), 'Biodiversity and stability in grasslands', *Nature*, **367**, 363–365.

Tilman, D., J. Knops, D. Wedin and P. Reich (2002), 'Experimental and observational studies of diversity, productivity, and stability', in A. Kinzig, S. Pacala and D. Tilman (eds), *Functional Consequences of Biodiversity: Empirical Progress and Theoretical Extensions*, Princeton, NJ, USA: Princeton University Press, pp. 42–70.

Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie and E. Siemann (1997a) 'The influence of functional diversity and composition on ecosystem processes', *Science*, **277**,1300–1302.

Tilman, D., C.L. Lehman and K.T. Thomson (1997b), 'Plant diversity and ecosystem productivity: theoretical considerations', *Proceedings of the National Academy of Sciences (USA)*, **94**, 1857–1861.

Tilman, D., P.B. Reich, J. Knops, D. Wedin, T. Mielke and C. Lehman (2001), 'Diversity and productivity in a longterm grassland experiment', *Science* **294**, 843–845.

Tilman, D., D. Wedin and J. Knops (1996), 'Productivity and sustainability influenced by biodiversity in grassland ecosystems', *Nature*, **379**, 718–720.

Tilman, D., C.L. Lehman and C.E. Bristow (1998), 'Diversity-stability relationships: statistical inevitability or ecological consequence?', *The American Naturalist*, **151**, 277–282.

Tisza, L. (1966), *Generalized Thermodynamics*, Cambridge, MA: MIT Press.

Torell, L.A., K.S. Lyon and E.B. Godfrey (1991), 'Long-run versus short-run planning horizons and the rangeland stocking rate decision', *American Journal of Agricultural Economics*, **73**(3), 795–807.

Townsend, K.N. (1992), 'Is the entropy law relevant to the economics of natural resource scarcity? Comment', *Journal of Environmental Economics and Management*, **23**, 96–100.

Trenbath, B.R. (1974), 'Biomass productivity of mixtures', *Advances in Agronomy*, **26**, 177–210.

Trenbath, B.R. (1999), 'Multispecies cropping systems in India', in E.C. Lefroy, R.J. Hobbs, M.H. O'Connor and J.S. Pate (eds), *Agriculture as a Mimic of Natural Systems*, Dordrecht: Kluwer, pp. 381–405.

Turner, R.K. (1999a), 'Environmental and ecological economics perspectives', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 1001–1033.

Turner, R.K. (1999b), 'The place of economic values in environmental valuation', in I.J. Bateman and K.G. Willis (eds), *Valuing Environmental Preferences. Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford, UK: Oxford University Press, pp. 17–41.

United Nations Environment Programme (UNEP) (2002), *Global Environment Outlook 3. Past Present and Future Perspectives*, London: Earthscan.

van der Heijden, M.G.A., J.N. Klironomos, M. Ursic, P. Moutoglis, R. Streitwolf-Engel, T. Boller, A. Wiemken and I.R. Sanders (1998), 'Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity', *Nature*, **396**, 69–72.

van der Heijden, M.G.A., J.N. Klironomos, M. Ursic, P. Moutoglis, R. Streitwolf-Engel, T. Boller, A. Wiemken and I.R. Sanders (1999), 'Sampling effect, a problem in biodiversity manipulation? A reply to David A. Wardle', *Oikos*, **87**, 408–410.

Van der Ploeg, F. and C. Withagen (1991), 'Pollution control and the Ramsey problem', *Environmental and Resource Economics*, **1**, 215–230.

Vandermeer, J.H. (1989), *The Ecology of Intercropping*, Cambridge, UK: Cambridge University Press.

Vandermeer, J.H. (1990), 'Intercropping', in C.R. Carrol, J.H. Vandermeer and P.M. Rosset (eds), *Agroecology*, New York, NY, USA: McGraw Hill, pp. 481–516.

Vane-Wright, R.I., C.J. Humphries and P.H. Williams (1991), 'What to preserve? Systematics and the agony of choice', *Biological Conservation*, **55**, 235–254

Varian, H.R. (1992), *Microeconomic Analysis*, 3rd edn., New York and London: Norton.

Vetter, S. (2005), 'Rangelands at equilibrium and non-equilibrium: Recent developments in the debate', *Journal of Arid Environments*, **62**, 321–341.

Vitousek, P.M. and D.U. Hooper (1993), 'Biological diversity and terrestrial ecosystem biogeochemistry', in E.-D. Schulze and H.A. Mooney (eds), *Biodiversity and Ecosystem Function*, Berlin, Heidelberg and New York: Springer, pp. 3–14.

Walker, B. (1992), 'Biodiversity and ecological redundancy', *Conservation Biology*, **6**, 18–23.

Walker, B., A. Kinzig and J. Langridge (1999), 'Plant attribute diversity, resilience, and ecosystem function: the nature and significance of dominant and minor species', *Ecosystems*, **2**, 95–113.

Wall, G. (1987), 'Exergy conversion in the Swedish society', *Resources and Energy*, **9**, 55–73.

Wall, G. (1990), 'Exergy conversion in the Japanese society', *Energy*, **15**, 435–444.

Wall, G., E. Sciubba and V. Naso (1994), 'Exergy use in the Italian society', *Energy*, **19**, 1267–1274.

Wardle, D.A. (1999), 'Is "sampling effect" a problem for experiments investigating biodiversity–ecosystem function relationships?', *Oikos*, **87**, 403–407.

Watson, R.T., V.H. Heywood et al. (eds) (1995a), *Global Biodiversity Assessment* (published for the United Nations Environment Programme), Cambridge: Cambridge University Press.

Watson, R.T., V.H. Heywood, I. Baste, B. Dias, R. Gámez, T. Janetos, W. Reid and G. Ruark (eds) (1995b), *Global Biodiversity Assessment. Summary for Policy-Makers* (published for the United Nations Environment Programme), Cambridge: Cambridge University Press.

Wätzold, F., M. Drechsler, C. Armstrong, S. Baumgärtner, V. Grimm, A. Huth, C. Perrings, H. Possingham, J. Shogren, A. Skonhoft, J. Verboom-Vasiljev and C. Wissel (forthcoming), 'Ecological-economic modelling for biodiversity management: Potential, pitfalls and prospects', *Conservation Biology*.

Weimann, J. and S. Hoffmann (2003), 'Brauchen wir eine ökonomische Bewertung von Biodiversität?', in *Messung und ökonomische Bewertung von Biodiversität: Mission impossible?*, Marburg: Metropolis, pp. 17–42.

Weimann, J., A. Hoffmann and S. Hoffmann (2003), 'Editorial', in *Messung und ökonomische Bewertung von Biodiversität: Mission impossible?*, Marburg: Metropolis, pp. 7–14.

Weinstein, M.C. and R.J. Zeckhauser (1974), 'Use patterns for depletable and recyclable resources', *Review of Economic Studies*, **41** *(Symposium on the Economics of Exhaustible Resources)*, 67–88.

Weikard, H.-P. (1998a), 'On the measurement of diversity', *Working Paper No. 9801*, Institute of Public Economics, University of Graz, Austria.

Weikard, H.-P. (1998b), Der Wert der Artenvielfalt: Eine methodische Herausforderung an die Ökonomik, *Zeitschrift für Umweltpolitik und Umweltrecht* **21**, 263–273.

Weikard, H.-P. (1999), *Wahlfreiheit für zukünftige Generationen*, Marburg: Metropolis.

Weikard, H.-P. (2002), 'Diversity functions and the value of biodiversity', *Land Economics*, **78**, 20–27.

Weikard, H.-P. (2003), 'On the quasi-option value of biodiversity and conservation', in J. Wesseler, H.-P. Weikard and R.D. Weaver (eds), *Risk and Uncertainty in Environmental and Resource Economics*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 23–37.

Weisbrod, B.A. (1964), 'Collective-consumption services of individual-consumption goods', *Quarterly Journal of Economics*, **78**, 471–477.

Weitzman, M. (1992), 'On diversity', *Quarterly Journal of Economics*, **107**, 363–406.

Weitzman, M.L. (1993), 'What to preserve? An application of diversity theory to crane conservation', *Quarterly Journal of Economics*, **108**, 157–183.

Weitzman, M. (1998), 'The Noah's Ark problem', *Econometrica*, **66**(6), 1279–1298.

Weitzman, M.L. (2000), 'Economic profitability versus ecological entropy', *Quarterly Journal of Economics*, **115**, 237–263.

Westoby, M., B.H. Walker and I. Noy-Meir (1989), 'Opportunistic management for rangelands not at equilibrium', *Journal of Range Management*, **42**, 266–274.

Whittaker, R.H. (1972), 'Evolution and measurement of species diversity', *Taxon*, **21**, 213–251.

Whitmore, T.C. and J.A. Sayer (eds) (1992), *Tropical Deforestation and Species Extinction*, London: Chapman & Hall.

Wicksell, K. (1893), *Über Wert, Kapital und Rente nach den neueren nationalökonomischen Theorien*, Jena: Fischer.

Wicksteed, P.H. (1992)[1894], *The Coordination of the Laws of Distribution*, first published in 1894, rev. edn., with an introduction by I. Steedman, Aldershot, UK and Brookfield, USA: Edward Elgar.

Widawsky, D. and S.D. Rozelle (1998), 'Varietal diversity and yield variability in Chinese rice production', in M. Smale (ed.), *Farmers, Gene Banks, and Crop Breeding. Economic Analyses of Diversity in Wheat, Maize, and Rice*, Boston: Kluwer, pp. 159–172.

Wiener, N. (1948), *Cybernetics or Control and Communication in the Animal and the Machine*, Cambridge, MA: MIT Press.

Williamson, A.G. (1993), 'The second law of thermodynamics and the economic process', *Ecological Economics*, **7**, 69–71.

Wilson, E.O. (ed.) (1988), *BioDiversity*, Washington, DC: National Academy Press.

Wilson, E.O. (1992), *The Diversity of Life*, New York: Norton.

Winkler, R. (2004), 'Time lagged accumulation of stock pollutants. Additively separable welfare functions reconsidered', *Discussion Paper* No. 408, Department of Economics, University of Heidelberg.

Winkler, R. (2005), 'Structural change with joint production of consumption and environmental pollution: A neo-Austrian approach', *Structural Change and Economic Dynamics*, **16**, 111–135.

Wirl, F. (2000), 'Optimal accumulation of pollution: Existence of limit cycles for the social optimum and the competitive equilibrium', *Journal of Economic Dynamics and Control*, **24**, 297–306.

Wirl, F. (2002), 'Stability and limit cycles in competitive equilibria subject to adjustment costs and dynamic spillovers', *Journal of Economic Dynamics and Control*, **26**, 375–398.

Wissel, C., T. Stephan and S.-H. Zaschke (1994), 'Modelling extinction and survival of small populations', in H. Remmert (ed.), *Minimum Animal Populations*, Berlin, Heidelberg and New York: Springer, pp. 67–103.

[WMPQ] *Wood Mackenzie's Pharma Quant* (1999), Edinburgh, UK.

Wootton, J.T. (1997), 'Estimates and tests of per capita interaction strength: diet, abundance, and impact of intertidally foraging birds', *Ecological Monographs*, **67**, 45–64.

World Bank (1997), *Expanding the Measure of Wealth. Indicators of Environmentally Sustainable Development*, Environmentally Sustainable Development Studies and Monography Series No. 17, Washington DC: World Bank.

Wu, J.J. and W.G. Boggess (1999), 'The optimal allocation of conservation funds', *Journal of Environmental Economics and Management*, **38**(3), 302–321.

Yachi, S. and M. Loreau (1999), 'Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis', *Proceedings of the National Academy of Sciences of the USA*, **96**, 1463–1468.

Young, J.T. (1991), 'Is the entropy law relevant to the economics of natural resource scarcity?', *Journal of Environmental Economics and Management*, **21**, 169–179.

Young, J.T. (1994), 'Entropy and natural resource scarcity: a reply to the critics', *Journal of Environmental Economics and Management*, **26**, 210–213.

Zeh, H.D. (2001), *The Physical Basis of the Direction of Time*, 4th edn., Berlin, Heidelberg and New York: Springer.

Zemansky, M.W. and R.H. Dittman (1997), *Heat and Thermodynamics: An Intermediate Textbook*, 7th edn., New York: McGraw-Hill.

Zhu, Y., H. Chen, J. Fan, Y. Wang, Y. Li, J. Chen, J.X. Fan, S. Yang, L. Hu, H. Leung, T.W. Mew, P.S. Teng, Z. Wang and C.C. Mundt (2000), 'Genetic diversity and disease control in rice', *Nature*, **406**, 718–722.