**The Dissertation Committee for David Moorfield Frank Certifies that this is the approved version of the following dissertation:**

# Values and Decisions in Biological Conservation

**Committee:**

________________________________

Sahotra Sarkar, Supervisor

________________________________

Jonathan Dancy

________________________________

Josh Dever

________________________________

Jim Dyer

________________________________

Cory Juhl

________________________________

Katie McShane

________________________________

Bryan Norton

# Values and Decisions in Biological Conservation

by

**David Moorfield Frank, B.A.**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

**December, 2012**

# Dedication

To Nora

## Acknowledgements

For his many insights and provocations, I am indebted to Sahotra Sarkar, without whom this dissertation would not exist. It has also benefitted greatly from conversations with dissertation committee members Josh Dever, Jonathan Dancy, Jim Dyer, Katie McShane, Bryan Norton, and Cory Juhl, prospectus committee members Hans Kamp and Dan Bonevac, Biodiversity and Biocultural Conservation Laboratory members Blake Sissel, Kelly Pierce, Andrew Tilker, and Chissa Rivaldi, graduate students at the University of Texas at Austin Nora Berenstain, Jeremy Evans, Kate Ritchie, Jeff Pote, Jeremy Colonna, Derek Anderson, Brian Cutter, and Jonathan Vanderhoek, and audiences at the University of Texas at Dallas, New York University, Cal Poly Pomona, and the University of Texas at Austin. Deborah Holle and Carl Schwope at the Balcones Canyonlands National Wildlife Refuge, Andrew French at the U.S. Fish and Wildlife Service, and biologists Larry Gilbert and Chuck Sexton provided crucial assistance. Finally, thanks to Nicholas Asher for teaching a fascinating course that forced me to learn game theory.

# Values and Decisions in Biological Conservation

David Moorfield Frank, Ph.D.

The University of Texas at Austin, 2012

Supervisor: Sahotra Sarkar

Conservation science applies research in the natural and social sciences to practical problems of nature conservation, thus presupposing various goals and values. This dissertation examines normative roles for the decision sciences in biological conservation. I am primarily concerned with two philosophical problems that arise in applications of the decision sciences to biological conservation problems: commensurability of multiple values and cooperation between multiple agents. I argue that models from decision analysis should be used to construct preferences over complex tradeoffs, and game theoretical models should be used to identify situations in which multiple agents pursuing their own interests cause outcomes that are worse for everyone. While these models allow values to be made explicit for decision-making, in other situations conservationists' goals and values are obscure. I discuss this distinct problem in the context of conservation biology, where the central concept of biodiversity is analyzed and shown to necessarily reflect the values of its users. The multiplicity of meanings of 'biodiversity' and measures of biological diversity raise risks for conservation biology and motivate multi-criteria approaches to conservation decision-making. Finally, the goals and values of conservation scientists and landscape managers may or may not reflect those of people who are affected by conservation policies. I argue that while decision science can aid in making values of various stakeholder groups

explicit, facilitating reflection and learning, it cannot resolve ethical dilemmas on its own without input from normative and applied ethics, particularly in identifying legitimate stakeholders and weighing multiple biological concerns against concerns for rights, welfare, and social justice.

## Table of Contents

# List of Tables

# List of Figures

# Chapter 0: Values and Decisions in Biological Conservation: *Introduction and Overview*

## 0.1. Philosophy of Conservation Science and Environmental Ethics

The chapters of this dissertation fall within the philosophy of conservation science and environmental ethics. In their recent introductory textbook on conservation science, Kareiva and Marvier (2011, 1) state:

> Conservation is both a scientific enterprise and a social movement that seeks to protect nature, including the Earth's animals, plants, and ecosystems. Conservation science applies principles from ecology, population genetics, economics, political science, and other natural and social sciences to manage and protect the natural world. Effective conservation requires a clear understanding of how people impact the planet and how they make decisions about their use of natural resources and their choice of lifestyle.

As characterized here, conservation science applies research in the natural and social sciences to practical problems of nature conservation, thus presupposing various goals and values. This dissertation will focus on problems of biological conservation, where the human goals and values at stake involve biological and ecological entities, like genes, organisms, populations, species, and ecological assemblages and communities.

Often the goals and values of conservation scientists are made explicit, for example in decision analyses of conservation problems where the values of human agents are specified. Part of this dissertation is concerned with philosophical issues that arise in prescriptive applications of the decision sciences to biological conservation problems, in particular the problems of the commensurability of multiple values (chapters 2-3, and 5) and cooperation amongst multiple agents (chapters 4-5).

Sometimes the goals and values implicit in biological conservation practice are obscure, whether to the producers or the consumers of conservation science, or both. Chapter 1 discusses this distinct problem in the context of conservation biology, where the central concept of biodiversity is analyzed and shown to necessarily reflect the values of its users.

Additionally, the goals and values of conservation scientists and managers may or may not reflect those of people who are affected by conservation policies. Chapter 5 discusses ethical aspects of this problem in the implementation of conservation, focusing on norms that ought to constrain tradeoffs and stakeholder analysis that go beyond the decision models discussed in chapters 2-4.

While Kareiva and Marvier do not explicitly include environmental ethics in their characterization of conservation *science*, it has been a primary goal of environmental ethics to articulate and defend an underlying justification for the goals and values that fall under the broad umbrella of "nature conservation" (Norton 1987, Sarkar 2005, Jamieson 2008). This dissertation will not contribute to this project. Rather its focus is more on the ethical aspects of practical problems. Thus I will assume throughout that investment of societal resources in biological conservation is justified to some extent. The chapters on normative applications of the decision sciences bring out ethical considerations characteristic of the problems of commensurability and cooperation. The final chapter deals with the ethics of commensurability and cooperation in the implementation of biological conservation by ethical reflection and examination of case studies.

## 0.2. 'Values' and 'Decisions'

In the title, and throughout, I use the terms 'values' and 'decisions.' Before I provide a more detailed roadmap of the chapters to come, I hope to briefly clarify these terms.

### 0.2.1. VALUES

There are many ways to define 'values' and measure people's values. Here I will not be engaged with first-order normative questions of what is valuable, but rather on the implications of what humans value, as measured by the psychological, behavioral, and social sciences. Thus values here are necessarily tied to human valuation, or valuing, the verb. In the broadest sense used here, to value something, whether a particular thing, a class of things, or something more abstract like friendship or the diversity of species in a biota, is to care about it positively, and to be disposed in one's attitudes and behaviors to exhibit this concern or interest.

Economists usually measure people's values by measuring their preferences. Preferences can be measured by observing consistent choice behavior. If I consistently choose $x$ over $y$, I am said to *prefer* $x$ to $y$, a statement which may be used to predict my behavior in similar circumstances. Cardinal *utilities* that include information about *how much* I prefer $x$ to $y$ may also be constructed from choices between uncertain gambles. Preferences may also be measured by simply asking or surveying people, or via more complicated techniques to construct a model of an agent's preferences, one of which will be surveyed in chapter 2. For expositions of the elementary decision and game theory that will be used in this dissertation, see Resnik (1987), Gintis (2009), and Keeney and Raiffa (1993).

Philosophers sometimes identify valuations with reflectively endorsed or "considered" preferences (Norton 1984) or preferences given full information (Gauthier 1986, Railton 1986). Others have defended views of *value* (that is, what is valuable) based on the idea that some valuations may be more justified than others in ways that go beyond requirements of full information or reflective endorsement. Anderson (1993) and others have argued that valuable things are the objects of *merited* or *appropriate* pro-attitudes, where the pro-attitude could be pleasure or desire but also respect, awe, love, etc. Here valuations may be quite distinct in their psychological properties depending on the pro-attitudes involved; the economic notion of preference clearly abstracts away from this kind of psychological detail. Other accounts also involve second-order normative judgments as to the reasonableness of valuations. For example, Scanlon (2000) proposed the so-called "buck passing" account of values such that to value something is to take oneself to have reasons to have certain attitudes.

In the chapters that follow I will usually use economic models of valuation as preference; however in chapter 2 I distinguish between revealed, elicited, and constructed preferences. There I will also rely on a normative claim that considered or reflectively endorsed preferences should be the basis of rational choice, particularly in high-stakes decisions.

There remain further questions about whether we may construct a philosophical or ethical account of unreasonable or pathological valuations or preferences. Such an account would be necessary to flesh out an account of value like Anderson's such that valuable things are appropriate objects of valuation. However we do not need to construct such a general account to make progress on the problems considered here.

### 0.2.2. DECISIONS

Identifying valuation with preference allows this often-amorphous concept to be specified with some precision and to play its standard role in decision theory. A decision is a situation where an agent must choose between multiple (>1) alternative courses of action. According to decision theory, agents choose (or, should choose) on the basis of their valuation of potential outcomes and the probabilities of various states of the world that, along with the agent's action, determine the outcome (Resnik 1987). Decisions may be complicated with multiple agents, leading to game theory (see chapter 4) and social choice theory, or with multiple criteria of evaluation, leading to multi-criteria approaches including multi-attribute utility theory (see chapter 2).

As Bermúdez (2009) points out, there are at least three “projects” for which decision theory has proven useful: guiding action, normatively assessing action, and explaining/predicting action. Economists have had some success using decision theoretic models to predict aggregate behavior in markets, as have behavioral ecologists and neuroeconomists modeling animal behavior (Glimcher 2011). Despite well-known anomalies and behavioral counterexamples to decision theoretic optimality, this research continues to identify situations in which we can expect humans and other animals to act as consistent maximizers. However, this dissertation will be concerned primarily with the other two projects, which are normative. Chapter 2 argues that decision theoretic tools of multi-attribute value theory can aid in guiding action by structuring our thinking about tradeoffs, and chapter 4 identifies a normative role for game theory.

Sarkar (2012a, ch. 4) has argued that decision theory can serve an important role bringing values and ethical considerations explicitly into the policy process. By specifying our valuations in decision theoretic analyses, we make those valuations and their consequences transparent. Besides the political benefits of transparency, this facilitates clear thinking (do you know what the consequences of your valuations are?) and reflection (are these *really* your considered valuations?). The idea that decision and game theory may serve as tools for ethical reflection will be a significant theme in the chapters to come. At the same time, decision theory does not provide a global theory of practical reason or a way to algorithmically bypass informal deliberation or political processes. They are tools, useful for particular purposes notwithstanding their limitations (Norton 2005, Norton and Noonan 2007).

Furthermore, as Hempel (1965) points out, the choice of decision models and decision rules also reflect the values of their users. To take a simple example, maximization of expected value or minimization of the risk of the worst outcome are two decision rules that will, in many cases, conflict in their recommendations. The decision-maker's "attitude toward risk," presumably dependent on all kinds of beliefs, valuations, and contextual factors, will non-trivially determine which decision rule is more appropriate. Chapter 2 discusses the philosophical and ethical assumptions necessary to use particular models in multi-criteria decision theory. Knowing *when* to use such techniques is as important as knowing how to use them.

## 0.3. Overview

Chapter 1 presents a critique of what has become the central concept of conservation biology, namely *biodiversity*. Consideration of the variety of definitions of 'biodiversity' and measures of biological diversity entails that any particular definition or measure used will depend on the user's goals and values. Empirical evidence that multiple measures of biological diversity can be non-concordant across landscapes and seascapes motivates multi-criteria approaches to biological conservation decisions. I call risks that arise in the context of conceptual engineering and operationalization for measurement "definitional risks," and compare them to risks arising from the acceptance of uncertain hypotheses (so-called "inductive risks," Hempel 1965, Douglas 2009). One axiomatic approach for constructing diversity functions over sets of pairwise differences serves as a case study of definitional risk. The chapter closes with the suggestion that the value of biological diversity may be dependent on facts about biological *composition*.

Chapter 2 moves to the philosophy of multi-criteria decision-making, focusing on a technique to construct a common scale of value (to "construct commensurability") when there are multiple values at stake. After presenting the technique, its scope, and limitations, I argue that constructing commensurability is a requirement of practical rationality for a certain class of decisions, namely those with high stakes and complex tradeoffs. This follows from the fact that it is implausible that decision-makers would even have considered preferences over complex tradeoffs without using such techniques, and the normative assumption mentioned above, that considered preferences are necessary for rational choice in high-stakes decisions. Chapter 3 presents a case study of constructing commensurability, the decision support system that the U.S. Fish and Wildlife Service uses to rank National Wildlife Refuges for budgeting purposes. I argue that the system has several flaws, most crucially that it does not take into account the marginal benefit of new acquisitions themselves.

Chapter 4 discusses a normative role for game theory in conservation contexts, specifically in identifying Pareto-inefficient Nash equilibria: cases where agents' pursuit of their own individual interests leaves everyone worse off than if they had cooperated or coordinated their behavior. Several case studies from the conservation biology literature are examined and modeled to show that such dilemmas are widespread. While several solutions to such dilemmas are available, particular attention is paid to the possibility of decentralized and community-based solutions. I appeal to behavioral game theory experiments that show that humans are often willing to enforce norms and cooperate, especially when there is opportunity for reciprocity over time.

Chapter 5 discusses ethical aspects of the problem of tradeoffs between multiple values and cooperation between multiple stakeholders, examining ethical assumptions that must hold for the kinds of decision theoretic arguments discussed in chapters 2-4 to have normative force. Several case studies reveal situations where decision science cannot provide ethical guidance without norms that delimit reasonable tradeoffs and identify legitimate stakeholders. Such approaches thus have significant limits when used in decisions that present ethical dilemmas.

# Chapter 1: Definitional Risk and 'Biodiversity': *Values in Conceptual Engineering at the Edge of Biology and Policy*

## 1.1. Introduction

Scientific measures and definitions used in applied contexts have non-epistemic consequences: their formulation is relevant to the utilities of decision-makers and stakeholders. Examples include measures and definitions of health and disease, poverty, economic welfare, the toxicity of regulated chemicals, and risk more generally.[1] This chapter illustrates how non-epistemic values[2] (hereafter, values) can affect scientific measures and definitions by examining their roles in defining 'biodiversity' and measuring biological diversity in conservation biology.

Douglas (2000, 2009) and others have argued that the existence of "inductive risks," non-epistemic risks associated with rejecting a true hypothesis or failing to reject a false hypothesis (type-I and type-II error, respectively), entails a necessary role for values in the appraisal of scientific hypotheses in applied contexts. This chapter makes a companion argument that the existence of *definitional risks*, non-epistemic risks associated with conceptual engineering and operationalization for measurement, entail distinct and equally necessary roles for values in applied science. Here I argue that the multiple ways of construing 'biodiversity' and measuring biological diversity raises the problem of definitional risk for conservation biology.

The chapter proceeds as follows. Section 2 distinguishes between inductive risks and definitional risks in applied science, shows why they are distinct, and gives examples

---

[1] Since risk is usually defined as the probability of disutility, or an expectation of disutility, how risk is measured and assessed will affect utilities when used in policy or decision-making.

[2] Epistemic or cognitive values include simplicity and explanatory/predictive power. Non-epistemic values (sometimes called "contextual values") include social, ethical, and aesthetic values. See Douglas (2009, ch. 5).

of both from conservation biology. Section 3 locates definitional risk in the context of defining 'biodiversity' and measuring biological diversity. Distinguishing concepts of richness, disparity, complementarity, evenness, and rarity across the biological hierarchy motivates pluralism about biodiversity concepts independent of pluralism in taxonomy. This shows that biodiversity cannot be captured by a single measure, raising the problem of definitional risk for single measures as well as composite indices that must make tradeoffs between several measures. Section 4 considers as a case study the axiomatic approach to constructing diversity functions from sets of pairwise differences of Weitzman (1992) and Gerber (2011). I argue that these axioms raise definitional risk in multiple ways and are consistent with many possible tradeoffs between richness and disparity. Section 5 closes by arguing that conservationists actually interested in biological *composition* should not appeal to the rhetoric of bio*diversity*.

## 1.2. Inductive and Definitional Risk

### 1.2.1. Two Kinds of Risks in Applied Science

The arguments from inductive and definitional risk are meant to show that values play necessary roles in applied science. The argument from inductive risk (1.2.2.) shows that values are necessary in setting the burden of proof for uncertain hypotheses in the face of possible disutility. That is, risks of error in an applied context play a role in determining how much evidence we require for a scientific claim in that context. The argument from definitional risk (1.2.3.) shows that values are necessary for determining our choices of a conventional definition (e.g. who counts as sick or poor, what counts as toxic, which species count as endangered, etc.) in the face of possible disutility. That is, risks in an applied context play a role in determining a term's conventional meaning. This section clarifies inductive and definitional risk, offers examples of both, and shows why they are distinct.

### 1.2.2. Inductive Risk

Douglas (2000, 2009) has argued, following Rudner (1953), Hempel (1965) and others, that if non-epistemic risks are associated with accepting a false hypothesis or rejecting a true hypothesis, values must play an indirect role in setting standards of acceptance and rejection. That is, in applying uncertain scientific claims to a decision (e.g. in making policy), the stakes of that decision should influence what we take to be *sufficient evidence* to believe a claim.[3] The argument is best stated using an example. Douglas (2000) considers the case of a hypothesis about the toxicity of dioxins and the decision to regulate these chemicals. Here the focus will be an example from conservation biology, namely uncertain estimates of *extinction probabilities* of a species of conservation concern derived from a population viability analysis (PVA) and the decision to stop or continue resource extraction in an area of the species' habitat.[4]

As Boyce (1992), Ludwig (1999), and others have argued, there is massive uncertainty associated with the results of PVAs, which can be due to scarce or poor data, sensitivity of results to model parameters that are difficult to estimate, and sensitivity of results to model assumptions that abstract away the complexities of actual populations and environments. Such complexities may include, among others, high variance in mortality and reproductive rates, demographic structure, spatial structure, multi-species interactions, and genetic effects (e.g. the effects of inbreeding). Demographic or

[3] This argument parallels arguments for contextualism about knowledge (see, e.g. Stanley 2005). For a history of this argument in the philosophy of science literature, see Douglas (2009).
[4] For reviews of population viability analysis, see Boyce (1992) and the papers in Beissinger and McCollough (2002). Beissinger and Westphal (1998) review the use of PVAs in endangered species management.

environmental "catastrophes" and other low-probability events that may raise or lower extinction probabilities are also difficult to incorporate into PVA models. Such uncertainties led Ludwig (1999) to argue that PVA results are often meaningless due to large confidence intervals, although Fieberg and Ellner (2000) argue that PVA results may be meaningful over short timescales. Here we will assume that the timescale is short relative to the time series data available, but of course there is still uncertainty associated with the estimated extinction probabilities.

Let us assume the statistical framework of classical hypothesis testing, and assume that our null hypothesis is that there is no difference between extinction probabilities in the *conservation scenario* in which the relevant area of habitat is left alone and the *extraction scenario* in which it is not. Such an analysis could be based on comparison of time series data from two different habitat patches, one that is intact and another where extraction takes place. An example of a similar analysis is found in Lindenmayer et al. (1993), where PVA is used to compare extinction probabilities for Leadbeater's possum (*Gymnobelideus leadbeateri*) when carrying capacity is or is not reduced by forestry practices.[5] We imagine that a one-sided statistical test tests whether greater extinction probability in the extraction scenario relative to the conservation scenario is merely due to chance. Our "significance level" $\alpha$ gives the upper threshold probability of our data were the null hypothesis true such that we reject the null hypothesis. Crucially, we assume that in this applied context, rejecting the null hypothesis constitutes a claim that extraction should be stopped.

Setting $\alpha$ relatively high will lower the burden of proof for rejecting the null hypothesis, creating a higher probability of a false positive (type-I error), claiming that

[5] See also Newman and Pilson's (1997) laboratory experiments on the plant *Clarkia pulchella*, where they established that decreased effective population size would lead to an increase in probability of extinction. Here they compared estimated extinction probabilities for two experimental populations.

extraction will raise extinction probability when it will not. Setting $\alpha$ relatively low will raise the burden of proof for rejecting the null hypothesis, creating a higher probability of a false negative (type-II error), claiming that extraction will not raise extinction probability when it will. Absent the prohibitively costly option of raising the statistical power of our experiment by increasing sample sizes, we must choose where to set the burden of proof. The claim is then that values *must* play a role in trading off type-I and type-II error in these cases. Not stopping extraction when it will be potentially disastrous for the endangered species risks species loss, while stopping extraction when it will not significantly affect extinction probability imposes an unnecessary cost on society.

Responses seeking to quarantine the role of values are easy to anticipate. One might claim, following Jeffrey's (1956) treatment, that we need not "accept" or "reject" this hypothesis at all, but merely hold it with degree of credence $p$, as the result of a Bayesian statistical analysis. This information can then be fed into a decision theoretic analysis, which would of course include values or utilities. One might respond to the classical hypothesis-testing example by claiming that as long as researchers report their p-values (the probability of the data were the null hypothesis true), they need not even engage in talk of statistical significance.

One response to these arguments admits that non-epistemic values are epistemically silent in the "pure" scientific process, as opposed to the applied scientific process that interacts with policy and decision-making. As Douglas (2009) shows, this was a widespread and popular view among postwar philosophers of science. For example, Jeffrey (1956) argued that it was not a scientist's job *qua* scientist to accept or reject hypotheses, but merely to hand over uninterpreted experimental results to decision-makers. While Douglas argues against this "value-free" ideal of science by appealing to ethical claims about the general responsibilities of scientists *qua* humans to consider the risks of error, it is not the purpose of this chapter to enter this particular ethical dispute.

The important point that both sides should be able to agree on is that the argument from inductive risk shows that values play an indispensible role in *applied* science. Applied scientists work in a context in which their results will potentially be used in making policy. Whether the scientists themselves take up setting the burden of proof, or the task is handed over to policymakers, it requires appealing to the values at stake in the relevant decision. It is worth reiterating in this context that probabilities derived from PVAs are associated with uncertainty that is difficult to quantify due to the fact that they rely on idealized models. Thus the suggestion that we simply believe the relevant hypothesis with some indeterminate credence $p$ and use this probability in a decision analysis oversimplifies the epistemic situation. The scientists or policymakers must make a decision about whether and how much to rely on the results of the PVA in the first place. This is an epistemic decision whose outcome will depend on the quality of the science *and* the stakes.

### 1.2.3. DEFINITIONAL RISK

The argument from *definitional* risk does not appeal to uncertainty in accepting or rejecting hypotheses, but rather to the conventionality of certain definitions and operationalizations for measurement in applied science. The argument is that if the use of different conventional applied scientific definitions or measures affects the utilities of decision-makers or stakeholders, these "definitional risks" should be taken into account in determining the definition or measure used in the applied context. The problem is particularly pronounced in medicine and psychiatry, as well as the social sciences. For example, governments may use economists' definitions of poverty in distributing the benefits of social programs.[6] I will argue below that it is also a deep problem in the case of defining 'biodiversity' for conservation biology, which is taken by practitioners to be a "crisis discipline" analogous to medicine (Soulé 1985; Sarkar 2002). Before moving to the case of 'biodiversity,' I present three examples of definitional risks in conservation biology: risks defining 'endangered', 'species range,' and 'evolutionarily significant unit' in assessing the conservation status of populations.

*'Endangered'*: Biologists apply the term 'endangered' to species at risk of extinction during a particular time period. The U.S. Endangered Species Act of 1973 defines 'endangered species' as "any species which is in danger of extinction throughout all or a significant portion of its range." (16 U.S.C. §1532 1973) Whether a particular species is in danger of extinction over a particular time frame is presumably a matter of fact that may depend on demographic, genetic, and environmental factors. However, the level of extinction risk that we deem acceptable for a particular species under the law is a normative matter, thus raising the problem of definitional risk for the term 'endangered,' whose determinate meaning must be established for any particular applied scientific context.

This is distinct from the problem of inductive risk, which would arise if we have already specified a determinate meaning for the term 'endangered' and then have to decide in conditions of uncertainty whether we have sufficient evidence to accept the hypothesis that a species is endangered. The problem of deductive risk may arise even in the (practically impossible) case where we have complete information on a species' risk of extinction. For example, say that know that a species has a 30% chance of extinction in the next 100 years. Whether that species should count as 'endangered' is a matter of

[6] The U.S. Federal Government's Department of Health and Human Services (HHS) lists 32 federal programs that use HHS poverty guidelines to determine eligibility to receive benefits (See U.S. Department of Health and Human Services 2012).

definitional convention that itself raises risks, for example that we may overcautiously waste resources protecting this species.

*'Species range'*: Generally biologists use the term 'species range' to refer to the geographic area where a species may be found. However, there are many ways 'range' can be construed: current range, historic range (at some point in history), native range (which would exclude captive or zoo populations), and potential range (suitable habitat that is not currently occupied but could potentially be colonized) are four examples. The Endangered Species Act uses the term in its definition of endangered species (see above), but does not specify which sense of 'range' is meant.

Vucetich et al. (2006, 1387) discuss the risks of equating range with *current* range: by ignoring the fact that the ranges of many species have been massively reduced by human activities, species with relatively low risk of extinction on very small current ranges may be delisted. While this is true, the other senses of range also carry definitional risks in this context. Using historic range may lead to many non-endangered species being listed, since some species with large stable populations have already experienced local extinction on large swaths of their historic range, for example the gray wolf (*Canis lupus*). Using historic range also raises the problem of specifying a reasonable timescale, again raising the problem of definitional risk in an applied context. The point is that since risk of extinction on current vs. historical ranges may vary widely, the definition of 'range' that is used in assessing risk of extinction will have downstream consequences for conservation resource allocation, raising the problem of definitional risk.

*'Evolutionarily significant units'*: A 1978 amendment to ESA allows listing "distinct population segments" of vertebrates, where these are local, geographically distinct populations that interbreed. The purpose of the amendment was to allow listing of locally endangered populations of species that are not globally endangered, for example

the bald eagle or the gray wolf in the lower 48 states, even though there were large, stable populations in Alaska. The concept of an "evolutionarily significant unit" within a species was introduced by Ryder (1986) to refer to groups within species that "represent significant adaptive variation." It was later used by the U.S. National Marine Fisheries Service (NMFS) to categorize distinct population segments of Pacific salmon for protection under ESA (Waples 1991). Waples argued that in order to qualify as a distinct population segment, a population should be an evolutionarily significant unit; that is, it should be a reproductively isolated, geographically distinct group with unique adaptations. This later became official NMFS policy (Pennock and Dimmick 1997). Several definitions of 'evolutionarily significant unit' were subsequently proposed in the biological literature, including definitions that appealed to morphological, phylogenetic, and genetic criteria of distinctness.

Pennock and Dimmick (1997) argue that the use of such definitions to identify distinct population segments for protection carries the risk that the original intent of the 1978 amendment would be lost. For example, if populations of bald eagles in the lower 48 states did not display unique adaptations, then they would not have counted as an evolutionarily significant unit. More importantly, while it is a matter of fact whether a particular population has adaptations that other conspecific populations do not, whether such adaptations should matter to conservation prioritization is a normative matter. Thus the applied use of any particular 'evolutionarily significant unit' concept, whether it relies on morphological, genetic, phylogenetic, or other criteria, implies definitional risk.

These three examples illustrate that definitional risks arise throughout conservation biology, which as an applied science must deal with both facts (what is the risk that the Bengal tiger will go extinct in the next 100 years?) and values (how should we prioritize Bengal tiger conservation?).[7]

Inductive risks and definitional risks are distinct. Inductive risks arise when evaluating an uncertain hypothesis is relevant to a decision, and we appeal to values in setting the burden of proof. Definitional risks arise at the stages of concept determination and operationalization for measurement (e.g. a rule for classifying a particular individual as poor or not-poor, or a species as endangered or not-endangered), thus affecting data gathering, modeling, and the formulation of hypotheses themselves. Here, values are consulted at this distinct stage in the scientific process.

While Douglas (2000) claims that systematic characterization of ambiguous data carries *inductive* risk (in her example, whether something should be counted as a tumor), I claim that they are better characterized as cases of definitional risk. Choosing a determinate extension or conventional meaning for a vague or general term raises distinct problems that should be kept separate from the argument from inductive risk, which relies on the role of values in setting the burden of proof or acceptance once a hypothesis is already formulated and tested. Similarly, in their wide-ranging discussion of what they call "methodological value judgments" in ecology, Shrader-Frechette and McCoy (1993) do not distinguish inductive risk from definitional risk.

---

[7] Elliott (2009) and Schiappa (1996) provide additional examples of definitional risks, although they do not use this terminology. Elliott focuses on cases from pollution research, arguing that linguistic and definitional choices surrounding the study of endocrine disruption, multiple chemical sensitivity, and chemical hormesis have downstream effects that are directly relevant to policy. For example, the U.S. Environmental Protection Agency defines an endocrine disruptor as any agent that "interferes" with the endocrine system, whereas the World Health Organization and agencies in Europe require demonstration of harm due to interference in the endocrine system. The latter definition clearly creates a higher burden of proof for whether a chemical should count as an endocrine disruptor, potentially affecting regulatory policy. Schiappa shows how the U.S. Federal definition of the ecological term 'wetland' was intentionally shifted for political reasons. George H.W. Bush had made a campaign promise that there would be "no net loss of wetlands" during his administration. In 1991 the administration produced a document bearing the name of the agencies charged with protecting wetlands that reduced by roughly a third the acreage that would count as wetlands in the lower 48 states.

I now turn to definitions of ‘biodiversity’ and measures of biological diversity in conservation biology, arguing that pluralism about biodiversity concepts raises the problem of definitional risk, particularly in the applied context of conservation prioritization.

## 1.3. Biodiversity Pluralism and Definitional Risk

### 1.3.1. Historical Context

The term ‘biodiversity’ was coined in the 1980s as a portmanteau of ‘biological diversity’ in an explicitly political context by scientists worried about the largely anthropogenic loss of species and ecosystems in the 20th century (Wilson 1988, Tackacs 1996).[8] However, ‘biological diversity’ as a theoretical term in the life sciences had existed at least since the 1950s (Magurran 2004), and human interest in natural variety is arguably as old as biology, or perhaps even as old as our species’ cognitive capacities for classification (Oksanen 2004). ‘Biodiversity’ has since become a term used widely by life scientists, environmental philosophers, policymakers, journalists, and environmental activists. The conservation of biodiversity as such, as a more general objective distinct from the conservation of particular species, ecosystems, or landscape features, has become the stated goal of many conservation organizations, as well as signatory nations to the 1992 Rio Summit’s Convention on Biological Diversity.[9]

### 1.3.2. Biological Diversities

Here, *biodiversity pluralism* is the view that there are multiple, incompatible ways of defining ‘biodiversity’ and measuring biological diversity. Such pluralism may arise in at least three places: (i) in biological taxonomy, where there exist multiple species

---

[8] See also Janzen (1986) and Soulé (1985).
[9] See Glowka et al. (1996).

concepts and strategies of classification (Maclaurin and Sterelny 2008, 32-33); (ii) in concepts of bio*diversity*, where there are distinct measures of variety, difference, and the biological diversity of a particular area; and (iii) in the mathematical operationalizations or formulations of these biodiversity concepts.

This section will focus primarily on type-(ii) pluralism, while the next section will focus on type-(ii) and type-(iii). While Koricheva and Siipi (2006) follow DeLong (1996) in stressing a strong distinction between conceptual definitions and operational measures of biodiversity, the two are clearly interrelated. Conceptualizations of biodiversity guide measurement strategies, and widely used metrics are often adopted as working definitions, for example species richness in Maclaurin and Sterelny's (2008) discussion of biodiversity. The discussion below traces this relationship by extracting a family of central concepts from the variety of metrics used by conservation biologists to measure biodiversity.

The most expansive—and, as Sarkar (2005) and others have pointed out, unhelpful—explicit definition of 'biodiversity' is that it is the variety of life at all levels of taxonomic and functional organization. This is particularly useless in the applied context since it is impossible to conserve all of life. Thus Sarkar and users of the systematic conservation planning framework (Margules and Sarkar 2007) take defining biodiversity in the applied context to involve identifying "constituents" of biodiversity: favored alleles, organisms, populations, species, or communities whose existence and persistence across space and time may be tracked for the purposes of conservation and management.[10] It should not be controversial that the selection of biodiversity constituents in this sense depends on our values: we would not target disease organisms

[10] Biodiversity *constituents* should be distinguished from *surrogates*, which are biotic or abiotic measures putatively correlated with units of conservation concern. See Margules and Sarkar (2007).

for conservation management, for example. However, the focus here is not on the selection of constituents, but rather the way the *diversity* of a biota is measured given some background set of biological or ecological "units" or systems.[11]

In this context, it is best to start with the practices of conservation biologists. Sarkar (2002) argued that we should take 'biodiversity' to be *implicitly* defined by the practices of conservation biologists, particularly the metrics and algorithms used by conservation biologists to prioritize areas for some variety of conservation management. Sarkar goes on to argue that these algorithms can be represented by a family of closely related concepts, captured by rarity and complementarity (a measure of the *new* units added to some background set of areas; see below). However he also admits that this view straightforwardly entails a kind of pluralism, since 'biodiversity' has been given countless definitions and operationalizations by conservation biologists and ecologists, and area prioritization algorithms used by planners are also many and varied.

Putting aside prioritization algorithms, the technical surveys of biodiversity measurement found in Gaston (1996), Magurran (2004), and Magurran and McGill (2011), reveal a wide variety of measures and indices, including simple species counts (usually called species richness or $\alpha$-diversity), relative abundance or evenness metrics, measures of commonness and rarity, indices of compositional differences between areas ($\beta$-diversity), and measures of functional, trait, and phylogenetic disparity. As mentioned

[11] This constituents-based approach is similar to the approach advocated by Maclaurin and Sterelny, where biodiversity is defined in terms of the number of biological *units* and the *differences* between them. While Maclaurin and Sterelny defend the common practice of using species as the core "unit" of biodiversity, units might in principle be taken at multiple scales. Some measure of difference or disparity between units (whether phylogenetic, functional/trait-based, etc.) is also necessary to capture the idea of bio*diversity*.

above, the pluralism runs deeper, since within most of these conceptual classes, many mathematical and statistical frameworks have been proposed.[12]

For this analysis of definitional risk, we may follow Sarkar's general approach, but focus more widely on the central concepts used by conservation biologists to define diversity. Given a particular taxonomic or functional unit of analysis (e.g. species or ecosystem), or a set of constituents, measuring the biodiversity of a particular area may take into account the following criteria, either individually or by combining measures in a composite index:

1. *Richness*: the number of units. Other things being equal, an area with more units (e.g. more distinct species) is more diverse than an area with fewer units.
2. *Disparity:* the differences between the units. Imagine two areas with the same species richness, but where the first area has many species that are closely related phylogenetically, while the second area has many species that are more phylogenetically disparate. Other things being equal, the second area is more biodiverse. While Faith (1992) and others advocate phylogenetic measures of disparity, other measures used include disparity of DNA sequence and morphological disparity, especially within a clade where a local "morphospace" may be constructed (Raup 1966; Maclaurin and Sterelny 2008, ch. 4; see the discussion of Neige 2003 below).

---

[12] For example, in a study of the concept of ecological diversity, which encompasses both the richness and relative abundance or evenness of species in an ecological community, Justus (2010) evaluates eleven distinct indices that take species richness and evenness into account. In their review of compositional similarity and β-diversity, Jost et al. (2011) list two incidence-based and eleven abundance-based similarity indices, which measure the similarity of two or more species assemblages. Velland et al. (2011) review indices of phylogenetic diversity, noting that conservation biologists and community ecologists have been using distinct types of measures with some overlap. For the first type of index (they list five in this class), a distinctness score is calculated for species in a superset phylogeny and then a function aggregates these scores for particular local subsets. The second type of index (of which they list four), distinctness scores depend only on these local phylogenies.

3. *Complementarity*: the number of new (distinct) units of an area relative to a background set. Other things being equal, an area with more distinct units adds more diversity to the total set than an area with less distinct units. Here related measures of the compositional difference between areas (β-diversity) are appropriate.[13]
4. *Evenness:* uniformity of the relative abundance of units. Imagine two areas with the same number of species, but in one area a single species dominates the ecosystem. Other things being equal, the area with a more *even* distribution (e.g. 40% species *A*, 30% species *B*, 30% species *C*) is more diverse than the area with a more skewed distribution (e.g. 90% species *A*, 9% species *B*, 1% species *C*).
5. *Rarity*: how rare the units are. Different kinds of rarity include abundance rarity (when there are few organisms of a species left), geographical rarity (endemics with limited range), and temporal rarity (a biological event that only happens rarely). An area with rare or endemic species is more diverse than one without, other things being equal.

### 1.3.3. Definitional Risks and 'Biodiversity'

The above list of central concepts used in measuring biodiversity for its conservation shows that biodiversity cannot be fully captured by a single measure, although single measures are often convenient to use in practice. Most importantly, notice the consistent use of the phrase "other things being equal" above. For any single measure of biodiversity $M(.)$, there may exist some other measure $N(.)$ such that, for areas $p$ and $q$, $M(p) > M(q)$, but $N(q) > N(p)$ such that we judge that *overall*, $q$ is more biodiverse than

[13] For a history of complementarity in designing conservation areas, see Sarkar (2012b).

$p$: $O(q) > O(p)$. For example, $p$ might have slightly higher species richness than $q$, but $q$ contains many more rare species than $p$.

Hughes et al. (2002) provide a more concrete example by showing that there is low concordance between species richness and centers of endemic fish and coral species in Indo-Pacific coral reefs. Analyzing a geographical database containing the ranges of 727 species of Indo-Pacific scleractinin corals and 1766 species of reef fishes, Hughes et al. looked at the relationship between species richness and endemism at 65 localities across the Indo-Pacific region. Whether endemism was defined as the lowest $10^{th}$ percentile of ranges in each major taxon, or in terms of an absolute area cutoff (they used 500,000 $km^2$, which identified 7% of corals and 28% of fishes as endemic), centers of high endemism and areas of high species richness were not found to be concordant for either corals or reef fishes. While they found strong correlation between overall species richness of corals and reef fishes, they found that lower diversity regions tended to have more endemics. This pattern is partly explained by the fact that the ranges of the most widespread corals and reef fishes tend to overlap near the equator, creating areas of high species richness where endemics are relatively less abundant.

In general, then, for any particular landscape or seascape, it would be implausible to assume without taxa-specific evidence that these multiple measures of diversity would be correlated. While area-specific biological and ecological knowledge can ameliorate the problem, definitional risk will arise whenever a single measure (e.g. species richness) is used in an applied context, since that measure may not capture all the aspects of biodiversity we care about (e.g. endemic species). Indeed, Hughes et al. conclude that the results of their study of corals and fishes suggest a “two-pronged” approach to reef conservation that takes both endemism and species richness into account.

Whatever underlying justifications for conserving biodiversity we accept, whether a duty to current or future generations of humans or to other species themselves, and/or the maximization of option value,[14] etc., there will be aspects of biodiversity that are more important than others. For example, efficiently maximizing option value may entail taking (at least) rarity, richness, and disparity into account. Thus a particular definition of 'biodiversity' used to prioritize areas to maximize option value that took only richness into account would entail significant definitional risk. In the case of Indo-Pacific coral reefs, we have evidence that a richness measure used for prioritization would not capture geographical rarity or endemism. Whether our measure *should* prioritize endemics depends on what our goals and values are in the applied context: this is the problem of definitional risk. To be justified, any use of such a one-dimensional measure in an applied context would have to be accompanied by evidence that all other important or valued properties *were*, at least roughly, equal.

Problems associated with using a single measure may motivate us to construct an *index* of biodiversity that takes multiple types of data into account, for example richness and disparity. Definitional risk arises here too, since defining quantitative indices involves making tradeoffs between the properties we wish to take into account. Making these tradeoffs will necessarily involve appealing to the values at stake in the applied context.

Morgan (2010) points out this problem for Maclaurin and Sterelny's units-and-differences approach by considering the case of trading off units (richness) and differences (disparity). To illustrate, Morgan poses the following thought experiment. Which set of numbers is more diverse, {2, 5, 9, 13} or {3, 25, 27}? Assume the natural

[14] Option value is the value we attach to conserving a resource (in this case, biological resources) so that we might retain the option of using it later (whether for scientific, aesthetic, economic, or other purposes). See Randall (1986).

measure of the disparity of a single set of numbers is just the sum of the positive differences between all the members. On this definition, the first set has 37 differences and four units; while the second set has 46 differences but only three units. If we judge the first more diverse, then one extra unit is "worth" nine differences, whereas if the second is more diverse, nine differences are "worth" more than one new unit. As Morgan (2010, 618) writes, "Whichever way we go involves a value judgment."

Morgan's argument is that if our overall assessments of biodiversity involve taking multiple criteria into account, any composite quantitative index intended to capture these judgments will involve trading off the multiple criteria. He does not present an argument that such tradeoffs will necessarily involve value judgments, but framing the problem in terms of definitional risk fills this gap.

Consider again the example of conserving biodiversity to maximize option value. Perhaps disparity would have to be weighted heavily relative to rarity and richness, since we judge that a more disparate but less rich biota retains more option value than a richer but less disparate biota. Here we must appeal to our values, in this case the underlying justification of biodiversity conservation, to make judgments about tradeoffs. In the example of Indo-Pacific coral reefs, any operational definition of 'biodiversity' for prioritization that would attempt to take species richness and endemism into account would be forced to trade these off.

Another empirical example of low concordance between multiple biodiversity criteria is provided by the dataset from Neige's (2003) study of the biogeography of Old World Sepiids or cuttlefish (Cephalopoda). Neige compiled occurrence data on 111 species of Sepiids from genera Metasepia, Sepia, and Sepiella, and arranged them into biogeographical units A-Q (see Figure 1) using as boundaries areas where multiple species' ranges meet. Data on 102 specimens' cuttlebone shape, representing 102 species, were used to construct a theoretical morphospace of possible cuttlebone geometries, using landmark-based geometrical morphometrics (Rohlf and Marcus 1993). "Landmarks" here were 15 loci on the ventral view of the cuttlebone that describe its shape, chosen because they were either junctions between major parts of the cuttlebone, centroids, or intersections of curves with a plane of symmetry. Thus the raw data consisted of 30 variables (X and Y coordinates of 15 landmarks) for each of the 102 species. Variations across landmark loci were represented in axes of a multidimensional "morphospace." Figure 2 plots species richness against "total variance," the sum of variances across axes in the morphospace, an index of morphological disparity.

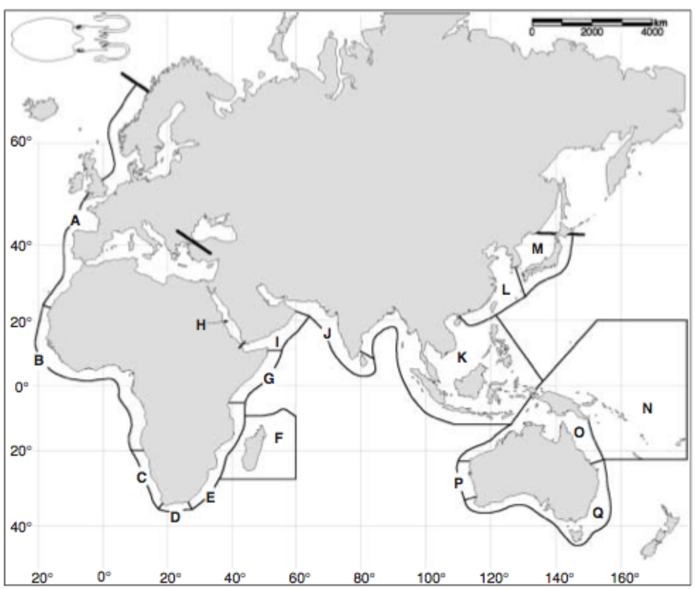



Figure 1.1. Geographical Regions of Sepiid Diversity and Disparity. From Neige (2003, 1126). Used with permission of Wiley-Blackwell.

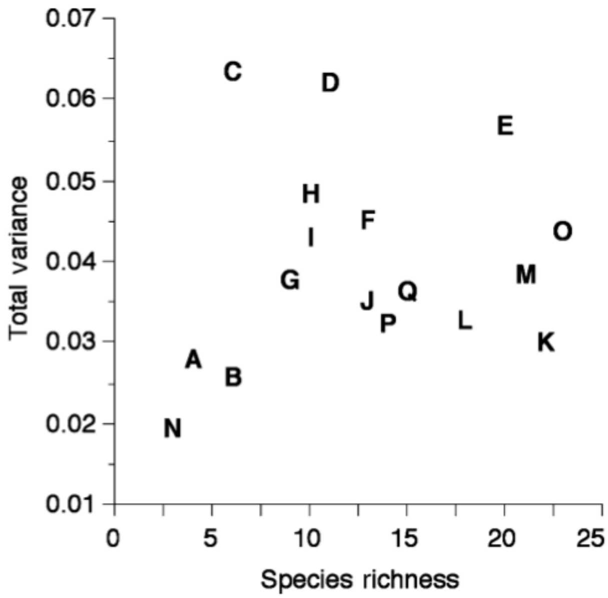



Figure 1.2. Species Richness vs. Total Morphological Variance of Sepiids from Biogeographical Regions A-Q in Fig. 1. From Neige (2003, 1134). Used with permission of Wiley-Blackwell.

As Neige notes, there is no simple relationship between cuttlefish species richness and this measure of morphological disparity: richness simply does not predict disparity. Area C off of the southwestern coast of Africa has the highest total variance but low species richness. Area O off the northeast coast of Australia (from the Tropic of Capricorn to Exmouth Gulf including the south coast of New Guinea) has the highest species richness but ranks below five other areas in disparity. Thus if we were to attempt to construct a quantitative index of overall Sepiid biodiversity of a area that took richness and morphological disparity into account, this index would have to trade off richness and disparity. How these tradeoffs should be structured depends on what we intend to do with this index, necessitating an appeal to our goals and values.

To summarize, this section has presented the following argument:

1. The practices of conservation biologists and our judgments about the relative biodiversity of areas entail pluralism about biodiversity, where the concepts of richness, complementarity, evenness, rarity, and disparity play important roles in defining ‘biodiversity’;
2. Any applied definition of ‘biodiversity’ (e.g. for prioritization) may take one or more of these central concepts into account;
    a. If only one measure (e.g. species richness) is used in an applied context, definitional risks arise, since other measures may contain information relevant to the values at stake;
    b. If multiple measures are used or aggregated into a composite index in an applied context, definitional risks arise, since making tradeoffs between multiple measures must appeal to the values at stake;
3. Therefore defining ‘biodiversity’ in an applied context carries definitional risk.

To reiterate an important point from the previous section, I have not here argued that it is in principle impossible to measure biological diversity or define ‘biodiversity’ without appealing to non-epistemic values. The discussion here leaves open the possibility of value-neutral theories of biodiversity as specified in some way. For example, the correct explanation(s) for the latitudinal gradient in species richness will not depend on the values of the investigators. However, the social and political contexts of the study of biodiversity, particularly the use of biodiversity concepts in policy and in allocating scarce conservation resources, necessitates care in locating roles for values in the *applied* scientific process. This discussion has located a role for values in defining ‘biodiversity’ and measuring biological diversity, since such definitions will entail definitional risk in applied contexts. The next section considers as a case study of definitional risk the axiomatic approach to constructing diversity functions over sets of pairwise differences originally proposed by Weitzman (1992).

## 1.4. Case Study: Weitzman's Axioms

This section considers in more detail the problem of definitional risk in defining biodiversity by considering axioms proposed by Weitzman (1992) for diversity functions that take sets of pairwise difference values and return a diversity value of the set. Definitional risk arises here due to the third type of pluralism described above, pluralism of mathematical operationalizations of biodiversity concepts. I first present the axioms and show how their acceptance leads to definitional risk. I also argue that many tradeoffs between units and differences are compatible with these axioms, a choice which also carries definitional risk in applied contexts.

Gerber (2011), following Weitzman's (1992) original treatment, takes the axioms below to be desiderata for diversity functions $V(.)$ that take sets $Q$ of pairwise distance values $d(i, j)$, representing the difference between units $i$ and $j$, and return a diversity value (where $Q \backslash i$ denotes the set $Q$ without unit $i$). Gerber and Weitzman take the units to be species, where disparity values could be generated via DNA hybridization experiments, comparison of DNA sequences, comparison of location in a theoretical morphospace, etc.[15] However, I present these axioms using the more general term 'units.'

1. *Montotonicity in units*: When a new unit is added to the set, diversity should increase. $\forall i, d(i, Q \backslash i) > 0 \rightarrow V(Q) > V(Q \backslash i)$.
2. *Twin property*: Diversity should not increase if the added unit is identical to one already in the set. $i, j \in Q$; $k, l \in E$; $E \cap Q = \varnothing$; and $E \cup Q$ is the total set of

[15] It should be noted that the use of any of these methods of measuring disparity requires justification. Whether there are non-arbitrary ways of quantifying, for example, morphological or genetic differences between species in distant clades, is not a question I will attempt to answer here.

units. If $d(i, k) = 0, d(i, j) = d(k, j)\ \forall j$ in $Q$ and $d(i, l) = d(k, l)\ \forall l$ in $E$, then $V(Q \cup k) = V(Q)$.

3. *Continuity in distances*: $V(Q)$ is continuous in distances. Sets of units $S$ and $S'$ contain $N$ units each; $S'$ is a function of $S$, such that $\psi(S) \equiv S'$. Define $\psi(Q) \equiv Q'$, both containing $N–k$ units. Then $\forall \varepsilon > 0, \exists \delta > 0$ such that if $\sum\sum|d(i, j) - d(\psi(i), \psi(j))| < \delta$, then $|V(Q)–V(Q')| < \varepsilon$.
4. *Monotonicity in distances*: $V(O)$ is monotonic in distances. For sets of units $S$ and $S'$, and $Q$ and $Q'$, with $N - k \geq 2$ and $\psi(.)$ as defined above. Then $d(\psi(i), \psi(j)) \geq d(i, j)\ \forall i, j \in$ Q and $i \neq j$, entail that $|V(Q')| \geq |V(Q)|$.
5. *Favor the most distantly related units*: If $d(1, i) > d(2, i)\ \forall i \in Q \backslash 1 \backslash 2$ and $1, 2 \in Q$, then $V(Q \backslash 2) > V(Q \backslash 1)$.

Acceptance of these axioms entails definitional risk, especially Axioms 1, 2, and 5. The first axiom states that diversity should increase when a new unit is added to the set. We may want this axiom to fail of our measure depending on the unit (e.g. species) under consideration: in the applied context, not all new species will lead to greater diversity. For example, if the added species is extremely common, or if we have some evidence that the added species may lead to a decline in the biodiversity of the region over the long term (consider the introduction of domesticated cats or an invasive plant), our judgments of overall diversity may fail this axiom. If monotonicity in units fails, then monotonicity in distances will also fail, since the introduction of a new (distinct) unit *i* entails that there exists a *j* such that $d(i,j) > 0$.

The second axiom entails definitional risk insofar as it formally rules out taking relative abundance data into account, since it states that adding the occurrence of an already-occurring species should not change the overall diversity value of the set. Common measures of ecological diversity (for example, those based on Shannon's entropy measure) that take richness and relative abundance into account are maximized as evenness is maximized.

According to Axiom 5, disparity maximizes diversity. Gerber (2011, 2279-2280) notes that Axiom 5 "suggests a value judgment pertaining to the optimization problem at hand." More than a suggestion, it *is* a value judgment when this measure is used in an applied context, namely the judgment that disparity is valuable. It is worth stressing here that disparity itself may be defined and measured in multiple ways, and the use of a particular disparity metric (e.g. phylogenetic disparity) will carry definitional risk, since it may or may not capture the property or properties we care about. In general, since these axioms apply to diversity functions over sets of pairwise differences, the use of other types of data (relative abundance or evenness data, or data on rarity, etc.) are formally ruled out from this type of analysis, entailing definitional risk when we have some valuation over variation in these properties.

Because the order units are input into the function can change a set's overall diversity value, Weitzman's function $V_W()$ that he proves satisfies these axioms is recursively defined as the maximum for all $i$ in $Q$ of $V_W(Q\backslash i) + d(i, Q\backslash i)$, which is unique when $V_W(i) = d_0 \ \forall i,$ where $d_0 = 0$ or any other constant. While Weitzman's function is plausible, *any* function (linear, exponential, logarithmic, hyperbolic, etc.) that is monotonic, continuous, and increasing in the addition of species and differences will satisfy these axioms. Since new units added to a set will have positive difference values paired with units already in the set, richness is also taken into account albeit indirectly. These various functions will represent different tradeoffs between adding new units and the differences between that new unit and the set. Insofar as a function satisfying these axioms does not satisfy other properties we may desire, for example decreasing marginal diversity of additional units added to a set, its use will entail definitional risk.

## 1.5. Biological Diversity and Biological Composition

### 1.5.1. 'BIODIVERSITY': SCIENCE OR POLITICS?

The scientists who originally introduced the term 'biodiversity' and its extremely catholic usage were almost certainly motivated more by political concerns than by scientific concerns. After all, if biodiversity is defined as the variety of life at all levels of organization, it becomes impossible to make *scientific* generalizations about biodiversity *in general*, as opposed to biological diversity as specified, for example in terms of species richness. Consider two descriptive studies, one that describes heterogeneity at a locus in a population of humans, and one that describes heterogeneity in species composition between habitat patches in tropical rainforest. According to the catholic usage (for example the consensus definition given in DeLong 1996),[16] these two studies are describing the *same thing*, or perhaps "aspects" of the same thing, namely biodiversity. This seems bizarre, given that the two studies are clearly looking at heterogeneity of very distinct systems at vastly different scales.

On the other hand, the introduction of the term could have had some heuristic value for biologists, since it focuses attention on heterogeneity as a central explanandum of the life sciences. Diversity properties have had limited success as explanantia, as the stalemate in the diversity-stability debate attests (Sarkar 2007). More often they figure as phenomena to be explained. For example, the increase in species diversity during the Cambrian explosion (500 mya) presents an explanatory challenge to macroevolutionists and paleontologists. In microevolutionary studies, the maintenance of genetic diversity

[16] DeLong's definition, based on consulting more than 80 published definitions: "Biodiversity is a state or attribute of a site or area and specifically refers to the variety within and among living organisms, assemblages of living organisms, biotic communities, and biotic processes, whether naturally occurring or modified by humans. Biodiversity can be measured in terms of genetic diversity and the identity and number of different types of species, assemblages of species, biotic communities, and biotic processes, and the amount (e.g., abundance, biomass, cover, rate) and structure of each. It can be observed and measured at any spatial scale ranging from microsites and habitat patches to the entire biosphere."

within populations became an explanatory problem after 20th century molecular techniques revealed more heterogeneity than many biologists had expected. As mentioned above, the latitudinal gradient in species richness also presents an interesting biogeographical explanandum.

The main *political* purpose of the introduction of 'biodiversity' was to focus attention on the loss of variety of species and ecosystems *in general*, as opposed to loss of particularly useful or charismatic species. By focusing on the value of maintaining diversity in general, specific justifications for conserving particular units, for example preventing the extinction of a migratory songbird in central Texas, become less important. Taking the conservation of biodiversity as a general goal also provides justification for conserving areas that probably contain undiscovered and unstudied species, tropical rainforests being a striking example.

One might argue that this strategy is problematic because no one believes it is either possible or desirable to conserve all species and ecosystems. Disease organisms and destructive "invasive" species are the best examples of species that we would not wish to conserve, at least *in situ*. But this argument will not work. Proponents of biodiversity as a value will respond that the conservation of biodiversity must be weighed against other social values such as economic welfare and public health. The analogous fact that public health should not be maximized at all costs does not vitiate the value of public health. However, there is still reason to doubt that biological conservation should be promoted *only* with reference to the value of biodiversity. This is because the value of biodiversity in some areas may be dependent on the value of the *particular* biological systems present in those areas. That is, this value would depend on the *biological composition* of those areas.

### 1.5.2. IS BIODIVERSITY'S VALUE DEPENDENT ON BIOLOGICAL COMPOSITION?

Biological composition refers to the particular biological systems (genes, organisms, species, ecosystems, etc.) that occur in a given area. (Two areas with the same species richness may obviously have very different biological compositions.) Landscape or seascape managers interested in maintaining ecosystem services like nutrient cycling or pollination, or sustaining high biomass yields, or protecting endangered or threatened species are more likely to be interested in the biological composition of an area as opposed to abstract measures of diversity like species richness. Pollination is only carried out by particular species. Biomass yields depend on the particular properties of the species being harvested. Anyone interested in preventing the extinction of a particular endangered species *in situ* is interested in maintaining a particular biological composition. Measures of diversity might be indirectly relevant to some of these goals. For example maintaining a variety of grasses might increase biomass yields in variable environmental conditions. However the *particular properties* of the varieties of grasses, in this case their tolerance of various environmental conditions, are what is causally relevant to maintaining biomass yields in this case. If it is the particular species present and their properties that determine management outcomes, a focus on biodiversity as opposed to biological composition is likely to be irrelevant at best, risky at worst.

Consider an analogy with the value of diversity in human institutions. The reason we may value diversity in a university setting, for example, is that people with diverse backgrounds and experiences can contribute to the culture of the university by exposing each other to different perspectives, preventing the potentially stultifying effects of intellectual and cultural homogeneity. However, we implicitly only consider cultures and viewpoints within a reasonable range: few would suggest that an undergraduate admissions committee always try to accept at least one neo-Nazi in order to maximize the diversity of viewpoints. Thus the value of diversity in human institutions also seems to be dependent on facts about the composition of such institutions. We are implicitly assuming that people from particular cultural backgrounds can make unique and valuable contributions to those institutions.

I am not making a general argument that maintaining heterogeneity or variety in some biological systems should not be one goal among many of conservationists. Indeed, the example of the value of undiscovered species, where we know very little about their particular properties, seems to show that we may also value biological diversity in ways that are only tenuously dependent on facts about biological composition. (The proponent of biological composition may argue that unless we know *something* about the undiscovered species, for example that it is a bird or amphibian, we have no reason to make any determination of its contribution to the biological value of an area.) Rather, I am arguing that when maintaining a particular biological composition is actually what is at stake, which seems to describe many cases of biological conservation, conservationists should not misleadingly appeal to the rhetoric of biodiversity.

## 1.6. Overview and Transition

This chapter has argued that definitional risk should be distinguished from inductive risk, and that definitional risk arises in defining 'biodiversity' and measuring biological diversity. Inductive risks arise in applied science when we have some intermediate degree of credence in a scientific hypothesis and we must set a burden of proof for that hypothesis when making a decision in the face of possible disutility. Definitional risks arise when definitions of terms used in applied science have downstream effects. I argued that definitional risk arises in defining 'biodiversity' due to its multifarious nature. I referred to empirical evidence that richness and endemism and richness and disparity can come apart, necessitating tradeoffs in composite indices of biodiversity. Finally, I considered a case study of axioms proposed for diversity functions, arguing that their acceptance may entail definitional risk, and argued that in some cases biological composition should be more important to conservationists than biological diversity.

The list of biodiversity criteria above, as well as the discussion of biological diversity and biological composition, both motivate multi-criteria approaches to biological conservation decisions. The next chapter discusses solutions to multi-criteria decision problems where a common scale is constructed to evaluate alternatives over multiple criteria.

# Chapter 2: Constructing Commensurability: *Tradeoffs, Practical Rationality, and Common Scales of Value*

> An emphasis on trade-offs in domesticated nature shifts the message of conservation from "No growth" or "Keep humans out" to "Be thoughtful about how humans conduct their lives and livelihoods." A key challenge for conservation science, then, is an accurate depiction of the many trade-offs that people face as they select and shape nature's future. (Karieva and Marvier, 2011, 22)

## 2.1. Introduction

### 2.1.1. Conservation Decisions with Multiple Values

Most important decisions in our personal and social lives have multiple values at stake. The last chapter motivated the use of multiple criteria in biological conservation decisions by showing that there are many conflicting ways to define 'biodiversity' and measure biological diversity. The definitions or measures used in a particular applied context depend on our goals and values, which may be many and varied. For example, while maintaining species richness might be desirable, additionally we may want to protect endemic species, where this goal may conflict with the goal of species richness. The problem of making decisions with multiple values at stake is even more difficult once we attempt to incorporate other societal goals, for example the minimization of economic opportunity costs.

One method decision analysts[17] have developed to begin structuring our thinking about multiple values for particular decisions is the *objectives hierarchy* (Keeney 1992, ch. 3). An objectives hierarchy shows relationships between fundamental objectives of a decision, sub-objectives, and measurable attributes that quantify the achievement of sub-

[17] *Decision analysis* is applied normative decision theory (Edwards and von Winterfeldt 1986), where the goal is to produce mathematical and computational procedures that aid in quality decision-making.

objectives. The achievement of sub-objectives helps achieve fundamental objectives. A fundamental objective is a goal or objective *O* such that there is no further answer to the question "why is *O* important?" that is directly relevant to that decision. Although fundamental objectives are of course open to further scrutiny, and may be justified with reference to other values or objectives, to be "fundamental" to a decision in this sense just means that such further justification is not relevant to the decision, and that the alternatives open to the decision-maker are all relevant to the achievement of that objective.[18]

Figure 1 below is an incomplete objectives hierarchy for a hypothetical conservation decision, including the fundamental objectives *maintain biodiversity* and *minimize cost*. Sub-objectives for maintaining biodiversity include maintaining species richness and protecting habitat for endemic species. One sub-objective for maintaining species richness is specified, namely to prevent the extinction of threatened species. Two sub-objectives for minimizing cost are also included. This objectives hierarchy is incomplete because measurable attributes that quantify the achievement of the sub-objectives have not been specified. (The next chapter will discuss desirable properties of attributes.)

[18] In the terminology of Keeney (1992, 2007), the fundamental objectives should be "controllable" (the alternatives influence whether the objective is achieved) and "essential" (all alternatives have such relevance).

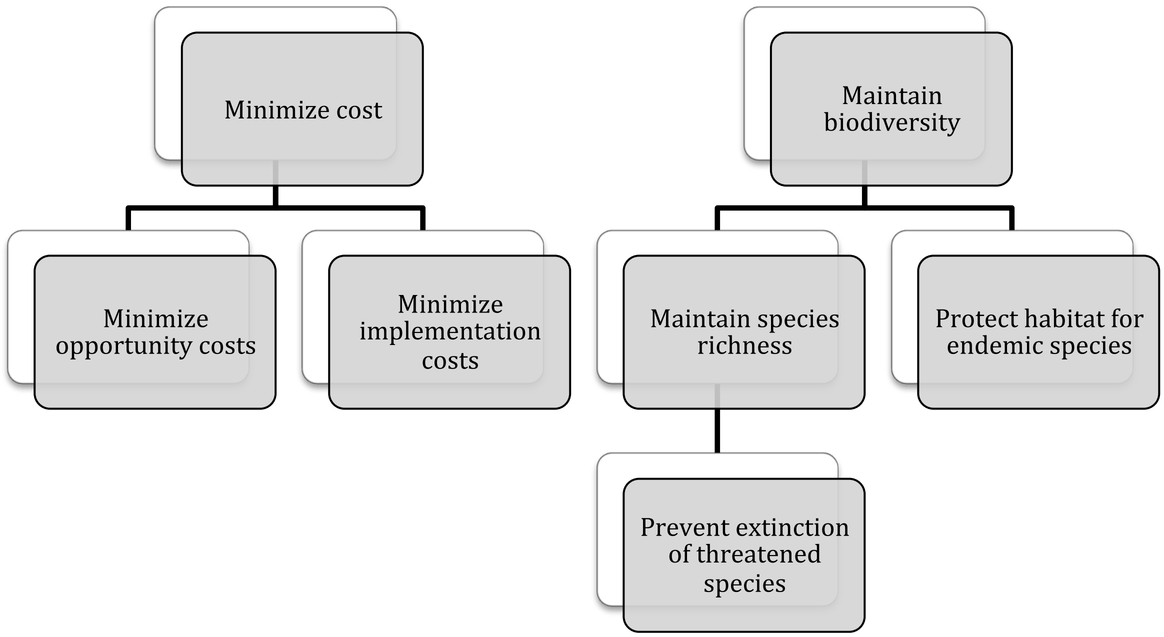



Figure 2.1. Incomplete Objectives Hierarchy for a Hypothetical Conservation Decision

As mentioned above, maintaining biodiversity and minimizing cost, although fundamental for this decision, could of course be justified in terms of *more* fundamental objectives that may be relevant to other decisions. This would take us into the philosophical domain of axiology or value theory, where the merits of various accounts of intrinsic value are debated (Schroeder 2012). (The language of values may be translated into the language of objectives by adding verbs like "to pursue," "to maximize," "to minimize," "to maintain," etc.) These deeper justifications for fundamental objectives could be offered by appealing to the goals of mainstream ethical theories: the pursuance of pleasure and avoidance of pain, or satisfaction of human preferences, or the achievement of other kinds of valuable states of affairs (as in versions of consequentialism), or acting according to impartial norms of rationality or reasonableness (as in versions of deontology), or the realization of virtues of character (as in virtue ethics), etc.

However, one feature of the objectives hierarchy approach in practice is that it allows one to avoid philosophical disputes about the justification of fundamental objectives in contexts where those disputes make no difference to the outcome of a decision. Thus an anthropocentrist and a non-anthropocentrist in environmental ethics could still agree that, for a particular policy decision, minimizing cost and maintaining biodiversity are both fundamental objectives, even though they disagree about the ultimate justification for those, or disagree about their relative importance.[19]

While objectives or goals should not be confused with values, fundamental objectives represent the decision-maker's values in the sense that for each fundamental objective, a distinct value is at stake (e.g. the value of biodiversity or economic welfare). Clarifying the values at stake for a decision by constructing an objectives hierarchy may aid in choosing between alternatives based on how those alternatives perform on attributes associated with sub-objectives. However, this raises the problem of how to weigh and make tradeoffs between the multiple values at stake, the subject of this chapter.

### 2.1.2. Multiple Values and Constructing Commensurability

In philosophical discussions of decision-making with multiple values, most authors have focused on the question of whether competing values (e.g. the value of biodiversity and the value of economic welfare) are *commensurable*, implying that they

[19] This is one way of putting Norton's "convergence hypothesis" (Minteer 2009), the idea that disagreement at the level of normative ethics or environmental ethics may disappear at the level of questions of sound environmental management.

may be traded off in a systematic way.[20] This chapter presents an argument that for a certain class of multi-value decisions, using decision analysis to construct a common scale of value, or to *construct commensurability*, is a requirement of practical rationality in those decisions. Since my focus is on commensurability and practical rationality in a particular class of decisions, I do not take a stand on the issue of whether *all* values (or, which values) are commensurable.[21]

The decision analytic technique for constructing common scales of value consistent with classical axiomatic decision theory, known as multi-attribute value theory (MAVT), has been available at least since the 1970s (Dyer and Sarin 1979).[22] This technique allows decision-makers to think quantitatively about the acceptability of various tradeoffs between values, which, at least in principle, may be easily measurable (e.g. quantities of harvested timber) or more difficult to measure (e.g. the aesthetic value of intact woods). Additionally, this technique does not necessitate the use of money as a numeraire. Preferences over complex tradeoffs are constructed in a dynamic process that involves modeling, learning, and reflection.

Discussion of this type of technique in the philosophical literature has been minimal, but has largely focused on its limitations.[23] Most of these limitations will be

---

[20] See, for example, discussions in Anderson (1993), Chang (1997), Griffin (1997), Chang (2001), Milgram (2002), Sarkar (2005), Aldred (2006), Trainor (2006), Kelly (2008), and Ellis (2008).

[21]For discussions of this issue, see especially the papers in Chang (1997). For a recent argument that all values are commensurable, see Kelly (2008). For arguments to the contrary, see Anderson (1993), Wiggins (1997), and Sarkar (2005).

[22] According to standard terminology, multi-attribute *utility* functions incorporate uncertainty, whereas multi-attribute value functions do not.

[23] For example, Anderson (1993, 49) discusses what she calls the "component value strategy" that "attempts to commensurate goods by representing the overall value of a good as an objective function of its component values." While she claims that this strategy may work in some multi-value decision problems (she focuses on examples of athletic scoring), she points out several problems, including the well-known result that this technique breaks down when alternatives can only be ordinally ranked on component values (Arrow and Raynaud 1986). See also Kagan's (1988) pessimistic discussion of additivity in ethics. For more technically sophisticated discussions, see Sarkar (2005) and Moffett and Sarkar (2006).

presented below, however I offer a more optimistic assessment of the technique, focusing on its ability to enhance practical rationality in the face of psychological biases, specifically tendencies to avoid tradeoff thinking and use simple decision rules with problematic implications.

#### 2.1.3. OUTLINE

The chapter proceeds as follows. Section 2 defines comparability and commensurability, and identifies the class of decision problems with multiple values under certainty where comparability of alternatives on each value may not suffice for rational decision-making. These decision problems are characterized by complex tradeoffs between values. Section 3 introduces the simplest linear technique as a solution, enumerating assumptions and idealizations that limit its applicability. Section 4 contrasts the decision analytic approach with two other ways of interpreting preferences: preferences revealed by choice behavior and preferences elicited by verbal behavior. This lays the groundwork for Section 5, which presents the argument for the practical imperative of constructing commensurability for high-stakes decisions where the assumptions are met.

### 2.2. Comparability and Commensurability in Multi-criteria Decisions

#### 2.2.1 COMPARABILITY

Here, a decision problem under certainty with multiple criteria will be characterized by:[24]

1. A finite set of alternatives $A$, where $a_i$ is alternative $i$;

[24] That a decision problem is under certainty could only be a plausible idealization. All decisions involve some form of risk or uncertainty. However we may abstract away from such considerations in cases where the uncertainty is sufficiently limited.

2. A finite set of criteria of evaluation $C$, where $c_j$ is criterion $j$;

3. Each criterion of evaluation $c_j$ inducing a *complete*, *transitive*, *weak* ordering ($\leq_j$) over the alternatives in $A$ such that:

3.i. $a_1 \leq_j a_2$ implies that $a_2$ is at least as good as $a_1$ on criterion $j$;

3.ii. if both $a_1 \leq_j a_2$ and $a_1 \geq_j a_2$ hold, this implies they are equally good on criterion $j$;

3.iii. if $a_1 \leq_j a_2$ but it is not the case that $a_1 \geq_j a_2$, then $a_2$ is strictly better on criterion $j$ ($a_1 >_j a_2$).

The third set of conditions defines *comparability* of alternatives on a criterion of evaluation. That is, a set of alternatives is comparable on a criterion of evaluation just in case the alternatives can be ranked ordinally by that criterion. The ordering is *weak* because ties are allowed, *complete* because all the alternatives are ranked, and *transitive* because $a_j \leq_j a_k$ and $a_k \leq_j a_m$ implies $a_j \leq_j a_m$.[25]

Comparability of alternatives on multiple criteria is often enough to eliminate some alternatives using the concept of *dominance*. An alternative dominates another just in case it is at least as good on all criteria and strictly better on at least one (Resnik 1987). For example, say a conservation organization is considering choosing one piece of land from an alternative set of two on three criteria of evaluation: species richness, ecological community uniqueness, and cost. The first piece of land is home to 150 species, but the ecological communities are not very unique, and the land is expensive. The second piece of land has roughly the same species richness and community uniqueness, but the land is less expensive. The second piece of land dominates the first on these three criteria of

[25] Comparability on a criterion should be contrasted with three kinds of incomparability: one in which it is *false* that any of the three possible value relations ('better than', 'equally good', 'worse than') hold (Raz 1997); one in which it is *true* that *all* or *more than one* of the relations hold (Seung and Bonevac 1992); and one in which it is *neither true nor false* that any of the relations hold.

evaluation, since it is equally good on the first two and better on the third. Assuming the criteria are exhaustive for the decision (there are no other relevant considerations), the second piece of land is clearly the choice that best reflects what the decision-maker cares about.

However, as the number of alternatives and criteria increase, dominance may only eliminate a few alternatives, and any alternative that is uniquely highest ranked on at least one criterion is *ipso facto* non-dominated.[26] For example, consider the three areas ranked ordinally (where 1 is the best, etc.) on the same criteria in Table 1.

| | *Species richness* | *Community uniqueness* | *Cost* |
|---|---|---|---|
| *Area A* | 1 | 2 | 3 |
| *Area B* | 2 | 3 | 1 |
| *Area C* | 3 | 1 | 2 |

Table 2.1. Difficult Decision Problem

*Area A* is the most species rich area, but is the most expensive and its community uniqueness value lies somewhere in between the other two areas. *Area B* is the cheapest, but has middling species richness and is home to the least unique ecological community. *Area C* has the most unique ecological community but is not species rich, and its cost lies

[26] Thus as Sarkar and Garson (2004) point out, the number of non-dominated solutions tends to increase with the number of criteria.

between the other two areas. Since each alternative is ranked first on at least one criterion, all of them are non-dominated. (Recall, for an alternative to dominate another, it has to be at least as good on all criteria and better on at least one. Since each alternative area is uniquely best on one criterion, in this particular example the first condition for dominance cannot hold.)

The next section characterizes in more detail why comparability of alternatives on criteria of evaluation, even along with comparability of the criteria themselves, may not be sufficient to make a choice that reflects the decision-maker's values.

### 2.2.2. Comparability of Criteria

In the decision problem posed above in Table 1, each area is non-dominated, so simple dominance analysis cannot be used to rule out any of the alternatives. One way to further enrich the basis for making decisions in these cases is to ordinally rank the criteria themselves in terms of importance. Thus a complete, transitive, weak ordering is induced on the criteria, implying that the criteria are comparable. This allows for use of the "Regime method" (Hinloopen et al. 1983), one of many proposed rules for multi-criteria decision-making that do not involve the construction of a common scale of value (Moffett and Sarkar 2006). The Regime method works as follows. For any two alternatives $a_1$ and $a_2$, let $K+$ be the set of criteria where $a_1 > a_2$ ($a_1$ is strictly preferred to $a_2$ on those criteria), and $K-$ the set of criteria where $a_2 > a_1$. Say that $a_1$ *outranks* $a_2$ if and only if $K+$ is non-empty and there exists an injective (one-to-one) function where each criterion in $K-$ is mapped to a more important (higher ranked) criterion in $K+$. This method yields as a solution the set of alternatives that are not outranked.

Assume that for the decision problem in Table 1, *Species richness* is more important to the decision-maker than *Cost*, which is more important than *Community*

*Uniqueness*. Comparing *Area A* and *Area B* with the Regime method results in *Area A* outranking *Area B*, since *Cost* is the only member of the set *K*-, which can be mapped to *Species richness* in the set *K*+. Comparing *Area B* and *Area C* results in *Area B* outranking *Area C*, since *Community Uniqueness* is the only member of the set *K*-, which can be mapped to *Cost* or *Species Richness* in the set *K*+. Comparing *Area A* and *Area C* results in neither outranking the other. Thus the only remaining alternative is *Area A*, which is outranked by neither *Area B* nor *Area C*. This example shows that Regime's outranking relation is not transitive in all cases, since in this case *Area A* outranks *Area B*, and *Area B* outranks *Area C*, but *Area A* does not outrank *Area C*.

The intransitivity of Regime's outranking relation is problematic if one takes transitivity to be a constraint on any rational ranking procedure for a single decision-maker, as classical decision theory assumes for the preference relation. The usual argument for this is that intransitive preferences can lead one to take a series of trades such that one is guaranteed to lose value. Say I prefer *A* to *B*, and *B* to *C*, but I am indifferent between *A* and *C*. If I start with *B*, someone could sell me *A*, then trade me for *C*, since I am indifferent between *A* and *C*. But then someone could sell me *B*, since I prefer *B* to *C*. If I start with *A*, someone could trade me for *C*, then sell me *B* and then *A* again. If I start with *C*, someone could sell me *B*, and then *A*, but then trade me for *C*. In all of these scenarios I lose value and end up with the alternative I started with.

Whatever one's philosophical view on the desirability of transitivity in a ranking relation,[27] more importantly Regime and other rules that do not involve constructing a common scale of value[28] do not allow us to think quantitatively about the acceptability of

---

[27] See Anand (1993) for a discussion. Rachels (1998) discusses purported counterexamples to the transitivity of "better than."

[28] See Moffett and Sarkar (2006) for a helpful review of many of these methods. Outranking methods such as PROMETHEE (Brans and Vincke 1985), where simple preference functions are constructed for each

tradeoffs between the values at stake. In this example, *Area C* outperforms *Area A* on both of the lower ranked (2nd and 3rd) criteria, but is outranked by *Area B*. Regime does not allow us to ask whether the superior performance of *Area C* on the two lower ranked criteria could compensate for its lower ranking on the most important criterion of *Species richness* in a pairwise comparison with *Area A*. Thus the choice of *Area A* may not reflect the values of the decision-maker.

Choosing any of these areas clearly involves trading off certain values against others. Simple cases like this one may only necessitate qualitative thinking about tradeoffs. That is, it may not be necessary to arrive at quantitative valuations on criteria for *Area A* and *Area C*, and then aggregate this information, to decide the compensation question. However, as decisions become more complex, with more criteria and more alternatives, and if the stakes are sufficiently high, we should be willing to spend time and energy analyzing tradeoffs quantitatively. Assuming we do wish to spend this time and energy, we would need to construct a *common scale* to decide how much of one kind of value we are willing to trade off against another kind of value.

To see more clearly why quantitatively analyzing tradeoffs might be desirable in this case, consider the following kind of argument. "*Species richness* is the most important criterion. Therefore we should choose the area that is highest ranked on species richness." This corresponds in outline to a decision rule where the alternative with the highest rank on the most important criterion is chosen. If there are ties on this criterion, the second-ranked criterion is used to break ties, and so on.[29] However there are many

criterion and then a partial or total preorder is defined over alternatives by aggregation via a "preference index," will not be discussed here. Unlike MAVT and MAUT, many outranking methods, including PROMETHEE, lack the axiomatic foundations of decision theory.

[29] Keeney and Raiffa (1993, pp. 77-78) call this a lexicographic ordering rule and argue on similar grounds that it is "rarely appropriate."

cases in which this rule will not result in decisions that reflect the values of the decision-maker, and others in which it won't aid in decision-making at all.

The first kind of case is one in which the range of species richness values is very small, whereas the range of, say, cost, varies widely. Thus, assuming the ordinal ranks on each criterion from Table 1, this decision rule entails choosing an area with an arbitrarily small positive difference in species richness value, but whose cost is also arbitrarily higher. Concretely, this decision rule would entail choosing an area with species richness of 151 that costs $1,000,000 over an area with species richness 150 that costs $10,000. Assuming the decision-maker cares about cost to some reasonable degree, this would likely be a choice that does not reflect the values of the decision-maker. This kind of case shows the importance of thinking quantitatively about tradeoffs, and that the "importance" of a criterion in a decision problem depends crucially on the *range* that the alternatives take on that criterion, a concept that will return below.

The second kind of case is one in which comparability of the criteria hold, but each criterion is deemed just as important as the others. Then this simple decision rule and Regime cannot eliminate any of the alternatives. New modeling assumptions will have to be introduced in order to make sense of these tradeoffs in a way that will aid decision-making.

### 2.2.3. Cardinal Value Functions and Commensurability Between Criteria

The last sections showed that comparability of alternatives on each criterion of evaluation *and* comparability of the criteria themselves might not be enough when decisions involve tradeoffs between values. These are cases in which, in order to be confident that our decision will reflect our values, we must consider two questions. Firstly, can we quantify *how much* one alternative outperforms another on each criterion?

This would involve translating our ordinal scales on each criterion to interval scales. Secondly, can we quantify *how important* each criterion is relative to the others, given the range of values taken by the alternatives on each criterion? This would involve translating our ordinal scale of the criteria themselves to an interval scale.[30]

We may answer the first question for alternatives $x$ and $y$ on some criterion of evaluation $c_j$ just in case a cardinal value function $v_j()$ can be constructed for that criterion such that (Dyer and Sarin 1979):

(1) $v_j(x) \leq v_j(y)$ iff $x \leq_j y$.

(2) $v_j(x) < v_j(y)$ iff $x \leq_j y$ and it is not the case that $y \leq_j x$.

(3) If $w$ is strictly preferred on criterion $j$ to $x$ and $y$ strictly preferred to $z$, then $\{w$ and $x\}$ is weakly preferred to $\{y$ and $z\}$ iff $v_j(w) - v_j(x) \leq v_j(y) - v_j(z)$.

This means that the alternatives can be ranked ordinally *and* that one can place the alternatives on a scale such that one can say *how much* better or worse one alternative is relative to the other alternatives.

If we performed this on all individual criteria, this may still not be sufficient to solve the difficult decision problem represented in Table 1. Even if we are able to say "how much more unique" the ecological community of *Area C* is relative to *Area B* and *Area A*, we still do not know how the consideration of ecological uniqueness is to be weighed against cost or species richness, given the ranges these variables take and how important differences within those ranges are to us. This requires commensurability of criteria of evaluation.

[30] Measurable value functions used here, in the sense of Dyer and Sarin (1979), are unique up to positive linear transformation, implying an interval scale of measurement. The use of ratio scales, although not formally ruled out by anything said here, imply the existence of a non-arbitrary zero point.

We may define commensurability of criteria by saying that criteria are commensurable if and only if an *overall* value function may be constructed that combines or aggregates the information contained in the value functions of individual criteria to produce a cardinal ranking of the alternative set. Decision analysts call this a *multi-attribute value function*. As above, 'attributes' are measurable quantities associated with criteria of evaluation. This hypothetical function would, for better or worse, render numerically precise the tradeoffs between values the decision-maker is willing to accept.

While numerical precision may be desirable for the sake of clarity or transparency, it may also raise several difficulties. The first may be due to judgments that appear arbitrary or insufficiently justified. For example, imagine being asked how much forest degradation, measured in parts per million of particulate pollution one would be willing to accept in exchange for a particular monthly discount in electricity costs due to the construction of a new coal plant. Making such a judgment in a reasonable and informed way would presumably require much information and reflection on the values at stake, including other relevant values like health impact. It is unclear whether anyone who is not an expert, or has not had much experience with this type of problem, would have well-formed preferences over such tradeoffs.

The second problem may be due to ethical considerations that limit the scope of tradeoff thinking. For example, if a forest is home to a group of humans who have a legitimate legal and ethical claim to the land, we may judge that their individual or collective rights should not be traded off against potential gains to others, perhaps within a particular range of potential gains or losses avoided.

As we will see in the next section, the construction of even the simplest multi-attribute value function involves accepting strong independence conditions and a potentially complicated procedure of construction.

## 2.3. Multi-attribute Value Theory (MAVT): the Construction of a Common Scale

### 2.3.1. The Additive Model

In the interest of focusing on the philosophical issues I only present the simplest kind of multi-attribute value function, namely one with an additive functional form, where the value of an alternative $a_i$ is a weighted sum of the values of $a_i$ on each criterion of evaluation $c_j$:

$$v(a_i) = \Sigma_j w_j v_j(a_i)$$

This says that the overall value of an alternative is a sum of the value of that alternative on each criterion of evaluation multiplied by a weight attached to that criterion. The weights $w_j$, which by convention sum to one, reflect the tradeoffs between criteria of evaluation the decision-maker is willing to make, given the variation present in the alternatives on each criterion. Different functional forms, for example a multiplicative model, will also imply different tradeoffs. The next two sections (3.2-3.3) will focus on the formal independence and ethical assumptions necessary to use the additive model in the first place, while section 3.4 will give an overview of the process of constructing the value functions and weights.

### 2.3.2. Ethical Assumptions: Trading, Trading Off, and Incommensurability

For such a model to usefully and accurately reflect the values of the decision-maker, they must be willing to make tradeoffs between the values at stake. Consider the following example, similar to the example of the ethical impermissibility of tradeoffs given above. Expand the decision problem represented by Table 1 to include the criterion of *extinction prevention*, and assume that only by prioritizing *Area A* can we protect a charismatic endangered species from extinction. Now assume that our decision-maker is a "no tradeoff absolutist" (Lemos 1994; Carlson 2001) about preventing extinction of charismatic species—we must prevent it at all costs. According to this no tradeoff absolutist, no world in which we allow this species to go extinct is more valuable than one in which we do not. Whether or not this is a reasonable view, this model cannot represent the values of our decision-maker, except trivially, by setting the weights of all other criteria to 0, and setting the extinction-prevention value of *Area A* to 1 and the extinction-prevention value of all other areas to < 1. The problem with this "trivial" formulation of the no-tradeoff view, however, is that it would presumably *not* reflect the decision-maker's values in other cases, where tradeoffs between, for example, cost of land and species richness *are* permissible. Furthermore, if each alternative involves species loss, the no-tradeoff absolutist is forced to confront tradeoffs behaviorally.[31]

There are several reasons why tradeoffs might be ethically problematic, and so commensurability between criteria compromised. However, it is important to distinguish the ethical acceptability of commensurability or tradeoffs in theory, as modeled by a multi-attribute value function, and substitutability or tradability in the market or a particular institutional or relational context. *Trading* certain goods may not be permitted in the open market (for example, we may think it is wrong to sell organs, sex, certain drugs or weapons, etc.), or within a certain institutional context (for example, we may think it is wrong to sell votes or good grades). Furthermore, certain goods may not be

[31] A more plausible model of the no tradeoff view involves modified lexicographic preferences with thresholds (Rosenberger et al. 2003; Georgescu-Roegen 1954), where tradeoffs are only permitted within a limited range of the relevant goods. For example, some minimum quantity of necessities (water, food, shelter, etc.) is necessary for life before any amount of those goods may be traded off. Similarly, there may be a threshold level of biodiversity such that our no-tradeoff absolutist allows tradeoffs above, but not below, that threshold. This threshold may be empirically tied to provision of necessary ecosystem services, but this would be controversial.

substitutable in the context of special relationships, for example a gift to my wife of a beautiful copy of her favorite book is not substitutable by its cash equivalent.

Such transactions or substitutions may weaken important institutions or relationships by undermining their main purposes or the motivations involved in sustaining them. If good grades could be bought, grades would cease to have whatever meaning and informational content they currently have in the institutional context of education. A gift of cash may be a bad gift to a good friend because it shows that one did not spend sufficient time or energy (or that one does not possess the information) to choose a particular gift.[32] More concretely, allowing trades between particular kinds of goods in the market may lower the quantity or quality of the goods provided. For example, Titmuss (1970) famously argued that allowing a market for blood could reduce the amount of blood available since it would crowd out people's other-regarding motivations to give blood.[33]

However, our judgments about inappropriate or unethical *trades* do not necessarily entail anything about the commensurability of the values at stake.[34] This is an important point that is missed by many philosophical commentators on the issue of commensurability, where there is an illicit slide from the inappropriateness of trading or market transactions to the general impermissibility of trading off.[35] Several examples should illustrate this point. Even though good grades should not be traded in a market, it

---

[32] Although see Waldfogel (1993) for an amusing argument that "deadweight loss" during Christmas could be avoided if some people gave cash instead of gifts.
[33]The empirical cogency of Titmuss's argument is contested, however see Mellström and Johanneson (2008). This study found that the crowding out effect for blood donations was significant for populations of women but not men, and that giving subjects the opportunity to give their compensation to charity counteracted the crowding-out effect. The idea that other-regarding motivations might be crowded out in certain settings, for example in markets, does have some support. See the discussion of motivational crowding out in Chapter 4.
[34] This point I owe to Dancy (personal communication).
[35] See, for example, Anderson (1993) and Sunstein (1997).

is not unethical for a decision-maker to trade off money and the prospect of better grades, for example by taking into account the cost of hiring a tutor. (This is not to say that someone who does so is consciously or unconsciously applying the kind of tradeoff reasoning used in MAVT.) Similarly, even though a thoughtful gift to my friend might not be substitutable for its cash equivalent, this does not entail it is inappropriate to trade off the intangible properties of gifts (for example, thoughtfulness) with cash, when making a purchasing decision as the gift giver.[36]

Certain limitations to tradeoff thinking, however, may derive from some of our deepest held ethical convictions whether or not they are reflected in legal restrictions on trading. Fiske and Tetlock (1997) review the anthropology and social psychology of "taboo tradeoffs," arguing that distinct norms govern distinct types of social relations or "spheres of justice" (Walzer 1983). Tradeoffs are more likely to be perceived as inappropriate where they transgress these norms. For example, explicit discussion of quantitative tradeoffs in kinship and intimate relations governed by norms of communal sharing are likely to be perceived as inappropriate. In the context of conservation, human groups' attachment to a place, with its particular geological, biological, historical, and/or cultural features, may be perceived by those groups as priceless.

It is arguably fundamental to our understanding of many human rights (for example, rights not to be enslaved, killed, tortured, or raped) that they may not be traded off, at least against certain kinds of goods (U.N. 1948). This kind of view is implicit in

[36] The controversial case of prostitution is interesting, since one may hold one of several plausible positions where the ethics of trading and trading off come apart. The first is that it is in principle not wrong to tradeoff sex for money, but that allowing sex markets create various kinds of harm (for example, harm to more permanent valuable relationships), thus these markets should be made illegal. Conversely, one might claim that it is in principle wrong to tradeoff sex for money (perhaps for virtue-theoretic reasons), but that criminalizing prostitution creates such harm (by creating an unregulated black market) that such transactions should be legal and regulated. There are, of course, two more logically possible positions where the ethics of trading and trading off do not come apart.

most deontological theories of ethics. Kant (2002) famously distinguished things with a *price* (fungible commodities tradable in the market) with things with *dignity* (persons). However, it may be permissible, indeed necessary or required, in contexts like war, to trade off between violations of rights themselves. For example, it might be permissible or required for a military commander to accept an alternative that includes certain rights-violations in order to avoid another alternative that includes even *more* rights-violations.

### 2.3.3. INDEPENDENCE ASSUMPTIONS

Besides potential ethical constraints, the additive model requires that the decision-maker's evaluations across criteria of evaluation are *additive independent* (Keeney 1992, 134-138) in that, (i) for all criteria of evaluation $j$ and $k$ where $j \neq k$ and for all alternatives $a_i$, $v_j(a_i)$ does not depend on $v_k(a_i)$; and (ii) for all criteria of evaluation $j$ and $k$ where $j \neq k$ and for all alternatives $a_i$ and $a_m$ where $i \neq m$, $v_j(a_i)$ - $v_j(a_m)$ does not depend on $v_k(a_i)$ or $v_k(a_m)$.[37] This says that, for all alternatives and all criteria, the decision-maker's evaluation of an alternative on a criterion (and the difference between two alternatives on a criterion) does not depend on the value of that (or those) alternative(s) on any other criterion. So tradeoff preferences cannot change depending on the levels of any of the individual value functions. For example, if the decision-maker is indifferent between two alternatives $a_1$ and $a_2$ on three criteria $j$, $k$, and $m$ where the values for alternative $a_1$ on $j$, $k$, and $m$ are $\{c,d,e\}$ and the values for $a_2$ are $\{c,d,e\}$, the decision-maker must also be indifferent between $a_1$ and $a_2$ when their values are $\{c,d,f\}$ and $\{c,d,f\}$, for all $f$. This holds for weak preference and strict preference as well.

Crucially, for any particular problem, additive independence is only required for the range of values obtained for the set of alternatives. This is important, since several

---

[37] See Dyer and Sarin (1979), Keeney (1992).

plausible lines of practical reasoning based on thresholds and holistic interaction effects violate additive independence. Consider a case where the level on one criterion of evaluation is so low, below some threshold of minimal acceptability, such that the values to the decision-maker on the other criteria of evaluation are significantly diminished. For example, if an area is degraded enough, it may not matter at all to a decision-maker interested in preserving biodiversity how cheap or politically feasible it is to acquire the land. This is a case where the evaluation of an alternative on one criterion, namely ecological quality, affects the evaluation of the alternative on other criteria, namely cost and political feasibility. If the ecological quality of the land were higher, cost and political feasibility would figure much more importantly in the decision-maker's practical reasoning, violating the independence conditions above.

Additive independence is violated in cases where the decision-maker's holistic evaluations over the alternatives depend non-additively on considerations that interact in complex ways. In his philosophical work on moral particularism, Dancy (2004, 2009) has defended the "holism of reasons," which holds, among other things, that the same reason may count in favor of an action in one circumstance but against it in another.[38] To borrow one of Dancy's examples, an action being against the law might in some cases be a reason to refrain from doing it whereas in some cases it might be a reason to do it, for example to protest the law. To use an example from conservation planning, species richness is often taken as a rough surrogate for the biodiversity value of a piece of land. But species richness may not matter as much relative to other considerations (in considering tradeoffs) in a circumstance where the species are neither endemic, rare, threatened, or endangered. Another example is the following. If local human settlements

[38] For another discussion of failures of additivity in moral reasoning, see Kagan (1988).

are considered threats to biodiversity due to sprawl or industrial or agricultural development, then conservation areas should be located as far away from human communities as possible. However, if local humans in certain settlements rely on natural resources as a buffer against poverty, perhaps the distance between these communities and the conservation areas should actually be minimized, not maximized.[39]

The decision analyst may respond to these failures of independence in several ways. One unlikely strategy would be to attempt to accommodate such reasoning by constructing more complicated value models with appropriate functional forms (e.g. multiplicative or polynomial) to accommodate the holistic and dependence effects in the judgments of the decision-maker. Unfortunately added mathematical complexity would come at the price of the scrutability, mathematical tractability, and ultimately the usefulness of the model. As the model becomes more and more complex, it becomes harder and harder for the user to understand and thus control the model.

The second, more plausible strategy is to restructure the problem such that the independence conditions can be said to hold, at least approximately, within the local context of the restructured problem. For example, additive independence may fail across the entire set of alternatives in the original formulation of the problem because some alternatives have a value on certain criteria below a minimum acceptable threshold. Simply removing these alternatives from the analysis may result in additive independence holding relative to the remaining alternatives. Additionally, if there are multiple measurable attributes that could be associated with a particular criterion, additive independence may fail for some attributes but not others. Dropping attributes for which additive independence fails may make the decision problem more tractable.

---

[39] See, e.g. Sarkar et al. (2011).

The restructuring strategy is perhaps the best response to the global holism of practical reasons: it is only in particular, local contexts that value is additive and all relevant values can be assumed to be additively independent. MAVT should not be interpreted as a global theory of practical reasoning about values, rather a locally useful way to think logically about permissible tradeoffs.[40] The chapter now turns to a brief introduction to the operational process of constructing an additive value function.

### 2.3.4. Constructing an Additive Value Function

We assume that the decision-maker, whether an individual or a group acting as a single agency, has reflected on the values at stake in the decision, enumerated them, and constructed an appropriate objectives hierarchy like the one at the beginning of this chapter. The decision-maker accepts the permissibility of tradeoffs in this context.

The decision-maker may ordinally rank each alternative on each criterion of evaluation, or directly move to the construction of value functions for each criterion of evaluation. This can be done in several ways. Perhaps the simplest way is to arbitrarily set the highest ranked alternative to 1, the lowest ranked alternative to 0, and ask the decision-maker to place the middle alternatives in-between, perhaps using a visual heuristic like a line segment. The analyst explains that these judgments should take into account the interval difference between the alternatives: if the value of $x$ is 1, the value of $y$ is .9, and the value of $z$ is .6, then $x$ is preferred to *y more than y* is preferred to $z$. The analyst may also show various shapes the value function could take (e.g. linear, convex,

[40] All scientific models introduce idealizations or approximations. It is an interesting question whether the kind of idealizations present in MAVT and other decision analytic methods can be placed in the taxonomies offered by philosophers of science, for example Weisberg (2007). This, however, is outside the scope of this chapter.

concave) and ask the decision-maker to identify one, potentially facilitating direct assessment or interpolation.[41]

At this stage the analyst should have verified whether the independence conditions associated with the additive model hold, by asking questions about whether the values assigned to alternatives on each criterion are independent of the other criteria.

Depending on the purpose of the analysis, the weights associated with each value function can be assessed directly using the following "swing" procedure, via a statistical regression procedure on holistic judgments, as well as other methods. In the swing procedure, the decision-maker is asked to imagine a situation in which all value functions had their lowest value. They are then asked on which criterion of evaluation would a "swing" from 0 to 1 be most preferable.

The swing weights crucially depend on the *range* of values taken by the alternatives on each criterion of evaluation. Even if species richness is intuitively the most "important" criterion for the purposes of planning for biodiversity conservation, the variation in species richness of the alternatives may be low enough within an acceptable range such that swinging from the worst to the best alternative is not as important as swinging from the worst to the best on, say, cost. The first-ranked criterion is then given a certain number of swing points, for example 100. The decision-maker then ranks how important each "swing" from 0 to 1 on each remaining criterion would be, relative to the first-ranked criterion, for example on a scale from 1 to 100, assuming none of the criteria should be given zero weight. The resulting scores are then normalized such that the

---

[41] Another way to directly assess the value of in-between alternatives is to use the von Neumann and Morgenstern (1944) method of gambles: if $x$ is the highest ranked alternative and $z$ is the lowest ranked alternative, have the decision-maker choose a probability $p$ such that, for alternative $y$, they are indifferent between $y$ for sure and a gamble that returns $x$ with probability $p$ and $z$ with probability $1$-$p$. If the value of $x$ is 1 and the value of $z$ is 0, then the value of $y$ is $p$. Other methods of indirect assessment or interpolation are available, for a review see Keeney and Raiffa (1993).

weight for a criterion is the swing points assigned to that criterion divided by the sum of all swing points.

Once the value functions and weights are constructed, the additive value function may be used to calculate the aggregate value of each alternative. This is often the beginning of a long process of analysis and refinement. The model is subject to thorough sensitivity analysis, where the parameters of the model are adjusted in various ways to see how the outcome is affected by changing those parameters. For example, the analyst may ask the decision-maker to re-consider each value function and provide a range of "acceptable" interval rankings. Broome's (1997) account of incommensurability as vagueness actually provides one underlying justification for sensitivity analysis: I may judge confidently that *Area A* has only a slightly more unique ecological community than *Area B*, whereas *Area C*'s community is *much* more unique than *Area A*'s. However, when asked to place these on an interval scale, I may find the following three value functions equally plausible (where $\{x,y,z\}$ gives the community uniqueness values assigned to *Area A*, *B*, and *C*, respectively, and the ranking on this criterion is *C*, *A*, *B*): $\{.25, 0, 1\}, \{.3, 0, 1\}, \{.35, 0, 1\}$.

Systematic sensitivity analysis should also be performed on the weights. For example, in their multi-criteria decision analysis of alternative conservation area networks in North-Central Namibia, Moffett et al. (2006) varied the weights for each criterion from 0 to 1.0, holding the relative weights of the other criteria constant, in order to see how variation in weights would affect the rank order of alternative conservation area networks. They also used simulation software to generate 10,000 random weights that preserved the ordinal ranking of criteria, and compared the resulting rankings of the alternative conservation area networks. The point is that a thorough sensitivity analysis should systematically vary value functions and other parameters including weights that may be affected by vagueness in judgments.

### 2.3.5. Review

So far, this chapter has (i) identified a class of decision problems where it may be desirable to construct a common scale of value; and (ii) introduced a simple mathematical model and a process of measurement that decision analysts have devised in order to do just that. The remainder of the chapter will offer a philosophical interpretation of this technique and consider the implications of this interpretation for practical rationality in decision contexts where the stakes are sufficiently high and tradeoffs must be considered.

## 2.4. Interpreting MAVT: the Construction of Tradeoff Preferences

### 2.4.1. Preferences: Revealed, Elicited, and Constructed

This section will distinguish three interpretations of preference, namely revealed preference, elicited preference, and constructed preference. The next section explains why one might want to construct a preference using decision analysis, specifically the presence of non-normative psychological biases.

The standard economic view of preferences abstracts away from preference-formation processes and the psychology and cognitive science of valuation. Preferences are modeled as static, complete and consistent ordinal or cardinal rankings over bundles of commodities or goods, or (more abstractly) states of affairs (Binmore 2008, Gintis 2009). Rankings over bundles of goods imply tradeoff preferences via the construction of indifference curves.[42] Both cardinal rankings over bundles and indifference curves imply the commensurability of the relevant goods.

While the theory of rational choice owes its historical origin to normative decision theory, most economists use rational choice models descriptively to predict behavior (usually in the aggregate), while decision analysts use the theory normatively to help their clients make better decisions (Edwards and von Winterfeldt 1986). According to the decision analysts, a decision by a client is *better* when it is informed and guided by the right process, which is identified with decision theory and its axiomatic apparatus of probability and utility (preference), often guided by heuristic aids like objectives hierarchies and influence diagrams.

Just as it is irrelevant to logicians how people *actually* reason theoretically, it is irrelevant to normative decision theorists how people actually reason practically. That people often affirm the consequent (infer $p$ from $q$ and *if $p$ then $q$*) is irrelevant to the logic of propositions; that people often commit themselves to intransitive preferences (ranking $a > b > c > a$) when asked to rank large sets of items is similarly irrelevant to the logic of preferences. However, the way people actually reason practically, and the psychological processes at work in valuation and preference, have important implications for *applying* decision theory. In particular, how the central notion of preference or utility is interpreted, and measured, depends on the application and use.

Economists pursuing descriptive applications usually rely on purely behavioral measures, or *revealed* preference. The idea is that studying stable consumption behavior on the assumption that individuals or firms are maximizing a utility function may

---

[42] An indifference curve in two dimensions would map the amounts $X$ and $Y$ of two goods such that the agent is indifferent between all bundles that lie along the curve (Keeney and Raiffa 1993). Each bundle on the indifference curve is non-dominated by the other bundles on the curve, and each bundle on the curve dominates bundles directly to the left and/or below.

facilitate prediction of future consumption behavior.[43] Other descriptive applications, for example contingent valuation surveys intended to measure humans' economic valuations over goods and services that lack a market price (ecosystem services, the "existence value" of an endangered species, etc.), involve the *elicitation* of preference via survey techniques (Carson 2011). These surveys usually ask subjects what they would be willing to pay, or willing to accept as compensation, in a hypothetical exchange scenario. The first distinction is thus between revealed preference, measured by choice behavior, and elicited preference, measured by linguistic (verbal or written) responses to particular prompts or questions.

While normative applications of decision analysis could in principle involve measures of revealed preference, they usually involve lengthy, complex processes of elicitation in conversation or dialogue, with significant interaction between the analyst and the client.[44] The decision analyst hopes to help the client arrive at stable judgments they would endorse upon reflection, which are furthermore checked for consistency and modeled mathematically. Contingent valuation survey methods, on the other hand, do not usually involve this kind of process. This is the distinction between *elicitation* and *construction* of preference: the former assumes (perhaps falsely) that the subject *has* a stable preference that can simply be measured by asking the right questions; the latter assumes that the subject has "basic values" but must arrive at a judgment of preference through a process of reflection and inference, with external aids to ensure consistency and stability. The process of construction *must* include elicitation as a component, but it also

---

[43] Revealed preferences are often interpreted, after Samuelson (1937), as being mere *descriptions* of choice behavior, as opposed to *explanations* based on psychological states, although this interpretation is controversial (Rosenberg 2005, Hausman 2000).

[44] See, for example, the dialogues found in Keeney and Raiffa (1993) or Edwards and von Winterfeldt (1986).

includes various checks on the consistency of subjects' responses, modeling and concomitant sensitivity analysis. This is because the constructivist assumes, consistent with behavioral research, that expression of preference can be highly contingent on the measurement procedure. Fischoff (1991, p. 835), describing this interpretation of value measurement associated with decision analysis, writes:

> [On this view,] people lack well-differentiated values for all but the most familiar of evaluation questions, about which they have had the chance, by trial, error, and rumination, to settle on stable values. In other cases, they must derive specific valuations from some basic values through an inferential process.

Decision analysts take the plausible line, defended here for the case of tradeoffs, that this "inferential process" ought to be compatible with our best normative theories of decision-making. So the process of construction should be set up to avoid the psychological effects that have been shown to result in systematic preference reversals and other non-normative anomalies. Some of these anomalies are reviewed in the next section.

### 2.4.2. Constructed Preferences and Non-normative Biases

The last section distinguished three ways in which preferences are measured and interpreted. The important distinction here is that between elicitation and construction. On the model of elicitation, stable preferences can be measured directly via linguistic behavior, responses to verbal questions or written surveys. However, depending on the circumstances, whether such stable preferences exist at all at the time the question is asked is unclear.[45] The literature in the behavioral sciences showing the sensitivity of elicited preference to various strategically irrelevant factors is large and growing (Slovic

[45] Clearly one may easily elicit trivial preferences, for example that people prefer more money to less, or health to disease.

1995, Lichtenstein and Slovic 2006).[46] For example, researchers have documented systematic preference reversals due to menu context (which items appear in a choice set) and framing effects (how a question is asked).

In a famous experiment, Tversky and Kahneman (1981) asked subjects to choose between two health treatments for 600 ill people, where option *A* will save 200 lives and option *B* has a 1/3 chance of saving all 600 and a 2/3 chance of saving 0 lives. They found that when framed in this way, a majority of subjects preferred *A*. However, when the same alternatives were framed in terms of *lives lost* (the first option loses 400 lives and the second has a 1/3 chance of losing 0 lives and a 2/3 chance of losing 600 lives), a majority of subjects preferred *B*. Similarly, Fischoff et al. (1980) found that a majority of subjects prefer a gamble that loses $200 with probability ¼ and returns nothing with probability ¾ to a sure loss of $50, but if the sure loss is framed as an "insurance premium," subjects generally prefer the sure loss.

The decision analytic response to these types of anomalies is to use our knowledge of normative decision theory in order to facilitate the construction of a stable, coherent, defensible, and reflectively endorsed preference structure (Payne et al. 1999). Rather than presenting the two treatment options in Tversky and Kahneman's case in terms of a single frame (or the two gambles in Fischoff's case), the decision analyst would present both frames to the subject, who would perhaps learn they are strategically equivalent.

In the case of considering tradeoffs, it has been shown that subjects perceive judgments about tradeoffs as difficult, and the degree of difficulty can be systematically predicted based on properties of the task, like the categories of the items to be traded off

[46] A factor is strategically irrelevant if it does not affect the decision theoretic representation of the decision problem. One example of a strategically irrelevant factor is the order in which alternatives appear on a menu.

and the perceived moral nature of the problem (Beattie and Barlas 2001). This difficulty (both cognitive and emotional) can result in the use of non-normative decision rules, or else people simply *avoid* thinking about tradeoffs altogether (Luce 1998, Fiske and Tetlock 1997). For example, there is some evidence that in multi-criteria decision problems, as the number of alternatives increases people often use a choice heuristic similar to the one considered and rejected in section 2.2, which chooses the item with the highest rank on the most important criterion (Payne 1976).

In their development of a provisional "building code" for constructed preferences, Payne et al. (1999) recommend the use of MAVT tools like the swing weight procedure in order to facilitate thinking about tradeoffs. They argue that the construction process should involve decomposing complex judgments about tradeoffs into simpler judgments, for example judgments about swing weights, and should include consistency checks and sensitivity analysis.

The position taken here, consistent with psychological findings and the practice of decision analysts, is that the preferences over tradeoffs measured by the MAVT process are not elicited or measured directly, but rather constructed. The process of construction includes elicitation as a component, but it also includes modeling and model sensitivity analysis, as well as the use of consistency checks to avoid the effects of frame-dependence and other systematic biases. This may seem a trivial point, but it has important normative implications. In particular, if having considered preferences over tradeoffs is a necessary condition for practical rationality in important decisions involving tradeoffs, and constructing commensurability is necessary to even *have* well defined, considered preferences over tradeoffs, then constructing commensurability is a rational requirement, at least for a certain class of decisions. The next section makes this argument in more detail.

## 2.5. Practical Rationality and Constructing Commensurability

### 2.5.1. Considered Preferences and Practical Rationality

One conception of practical rationality implicit in decision theory is that rational choice maximizes expected utility; another, which assumes only comparability, is that rational agents choose an alternative that is at least as preferred as all other alternatives.[47] However, many philosophers, for example Norton (1984), Gauthier (1986), and Railton (1986) among others, have been attracted to the idea that a rational choice can only be made in light of "considered" preferences, or preferences that the agent would endorse upon reflection, perhaps given the best available information and integration with their other beliefs and values.

Motivating this view is the fact that some agents act on desires that are formed under circumstances that are far from ideal, for example, involving highly addictive drugs, or complex and confusing circumstances. While such actions might be rationalizable according to *some* preference order, they are arguably not rational choices, since they do not reflect the agent's considered preferences. Here the sense of 'rational' has shifted from implying simple standards of consistency to implying internal standards of reasonableness. While one could develop a theory of rationality for considered preferences themselves that discriminates preferences as unreasonable based on their content (e.g. ruling out various self-destructive preferences), here I am only interested in minimal standards required for an agent to have considered preferences over a set of alternatives in the first place. Minimally, an agent should be able to articulate their preferences by considering their valuations over the alternatives, and would endorse those

---

[47] As mentioned above, this notion of practical rationality may only be locally useful.

preferences upon reflection and further consideration of the consequences of those preferences. In the case of tradeoffs, the claim is that agents facing decisions with complex tradeoffs who have not considered the tradeoffs between the values at stake would not have the basis to make a rational decision.

### 2.5.2. Constructed Preferences and Constructing Commensurability

Say that the constructivist interpretation holds for some agent $A$, set of alternatives $O$, preference $P_A$, and process of construction $C$ if and only if $A$ would not have preference $P_A$ over the alternatives in $O$ without engaging in process of construction $C$. A preference $P_A$ is an ordinal or cardinal ranking of alternatives in the set $O$. Processes of construction are meant to exclude the simplest elicitation procedures—we assume there is sufficient complexity that it is necessary to model the agent's preferences. Thus the constructivist interpretation would *not* hold for my preference for more money to less money, or health to disease.

Put simply, the constructivist interpretation of a preference holds when an agent would not have that preference without engaging in the process of construction. The psychological literature cited above supports this claim in the case of complex tradeoffs: people will usually lack well-defined preferences (let alone considered preferences) over alternatives involving complex value tradeoffs unless they are forced to consider them explicitly, for example in a process like MAVT. This is because people are likely to avoid tradeoff thinking due to its cognitive complexity and emotional difficulty, and when they do consider decisions with complex tradeoffs they are likely to use non-normative decision rules.[48]

[48] See especially the papers in Lichtenstein and Slovic (2006). From the perspective of a practicing decision analyst, Keeney (2002) discusses common mistakes in making tradeoffs, including assessing value tradeoffs independent of the range of consequences and using threshold rules for eliminating alternatives without considering how those alternatives perform on other criteria.

Recall that MAVT and MAUT (multi-attribute utility theory) methods are consistent with classical normative decision theory (Dyer and Sarin 1979, Keeney and Raiffa 1993), and the process of construction employed by decision analysts is meant to avoid psychological pitfalls while allowing for reflection (Payne et al. 1999). So the constructivist interpretation will hold for preferences constructed using MAVT, and furthermore will be compatible with our best normative theories of decision-making.[49] When used correctly, these methods lead to a satisfaction of the minimal requirements for considered preferences given above, namely that an agent should be able to articulate their preferences and would endorse those preferences upon reflection and consideration of their consequences.

Putting the crucial claims together, the argument for the requirement of constructing commensurability goes like this:

(1) *Empirical claim*: In important decisions with complex tradeoffs, we would not have considered preferences over tradeoffs without constructing commensurability.

(2) *Normative claim*: Having considered preferences over tradeoffs is a necessary condition for practical rationality in important decisions involving complex tradeoffs.

(3) *Conclusion*: Practical rationality in important decisions involving complex tradeoffs requires constructing commensurability.

The first claim is supported by empirical psychology and the experience of decision analysts. The second claim is philosophical, and the best argument for it is

[49] The classic argument for the view that the axiomatic foundations of decision theory are constraints of rationality is the (diachronic or synchronic) Dutch book argument (Vineberg 2011). Different versions of this argument show that if one's preferences or probability judgments are not consistent one can be offered bets such that one will take them and be guaranteed to lose value. While the cogency of this argument is contested, alternately we may assume that decision-makers have a preference for consistency.

simply to consider the alternative. Say there is a high-stakes decision with complex tradeoffs. Imagine a decision-maker who has simply not considered the relevant tradeoffs. It seems implausible that they could make a rational decision, especially in light of the empirical evidence on human judgment and decision-making cited above. The decision-maker might rely on their holistic judgment, but there is empirical support for the claim that in sufficiently complex cases, such judgments would likely result in a decision that would have counter-normative implications and would not be endorsed upon reflection.

While this argument focuses on a necessary condition for synchronic practical rationality, it should also be noted that use of such decision tools could enhance diachronic practical rationality for adaptive management over time (Holling 1978; Norton 2005). Wasting time and energy constructing new value models for similar decisions could be avoided, while common scales may be tuned or more fundamentally changed over time as new circumstances arise and priorities shift or new values must be incorporated. Use of such a system also facilitates post hoc assessment of decisions, since the values at stake and their tradeoffs are made explicit and transparent.

Finally, the decisions to which the argument applies are identified as “important.” As the stakes of a decision rise, it makes sense to invest more time and energy into the decision-making process itself. One only ought to apply normative decision analytic techniques in cases where decisions are sufficiently complex, unfamiliar, or otherwise cognitively demanding, *and* where the stakes are sufficiently high. For example, one need not apply decision theory to help someone make better decisions about weekly grocery purchases, or other everyday transactions where people’s preferences have been settled after processes of learning. Keeney (2004) makes this point by imagining a set of 10,000 decisions, where 7,000 have small consequences and 2,000 are “no-brainers.” Thus only

1,000 are worth thinking about at all, and of these, many *can* be resolved by clear, informal, holistic thinking. Of the decisions that receive systematic thought, some can be resolved by clear description of consequences or objectives. Others necessitate clear, systematic thinking about tradeoffs or uncertainty.

## 2.6. Conclusion and Transition

This chapter has presented a philosophical rationale for a decision analytic procedure for considering tradeoffs (multi-attribute value theory) by coupling a plausible empirical claim (that people lack preferences over complex tradeoffs but they may be constructed) with a plausible normative claim (that considered preferences are necessary conditions for practical rationality in choice). The next chapter will consider the application of multi-criteria decision analysis in biological conservation, focusing on the Land Acquisition Priority System (LAPS) developed by the U.S. Fish and Wildlife Service to prioritize National Wildlife Refuges for budgeting purposes.

# Chapter 3: The Land Acquisition Priority System (LAPS): *A Case Study in Constructing Commensurability for Conservation Prioritization*

## 3.1. Introduction and Overview

The last chapter was concerned with the philosophy of decisions over alternatives with multiple criteria of evaluation. The main argument was that for a certain class of complex, multi-value decisions with high stakes, constructing commensurability is a requirement of practical rationality. This chapter develops a case study of the construction of commensurability for conservation prioritization, focusing on the Land Acquisition Priority System (LAPS) used by the U.S. Fish and Wildlife Service (FWS) to rank National Wildlife Refuges (NWRs) for budgeting purposes. LAPS is explicitly multi-criterial: an NWR's LAPS score includes information about its fisheries and aquatic resources, endangered and threatened species, efforts toward bird and ecosystem conservation, and variables related to the status of the project on the landscape. However I will argue that it has several flaws and could be improved.

Below I qualitatively and quantitatively identify and the values in the system and, along with more specific criticisms, advance three general criticisms of LAPS: (1) *no complementarity*: it does not take into account the marginal value of new acquisitions, only the value of each NWR as a whole; (2) *well-roundedness*: it prioritizes NWRs that are "well-rounded" potentially at the expense of important projects with more narrow focus; and (3) *no social scientific data*: it does not sufficiently take into account information about surrounding human communities, for example the existence of programs that involve the public, or data on threats due to development or local hostility. The first two problems relate to the kind of decision analysis performed by LAPS, whose purpose is to create a prioritized list or "portfolio" of refuges to aid funding decisions, not to choose a single best alternative.[50] This raises problems with using an additive evaluation model like LAPS, discussed in section 4 below.

The chapter proceeds as follows. Section 2 introduces in outline the LAPS decision support system and FWS's statutory responsibilities, and presents the methodology of the chapter. Section 3 breaks down each criterion in more detail, constructs an implicit objectives hierarchy for each criterion, and presents specific criticisms of each criterion. Section 4 presents the more general criticisms, and section 5 connects the concerns of the last chapter to LAPS and transitions to the next chapter.

## 3.2. LAPS: Prioritizing National Wildlife Refuges

### 3.2.1. BACKGROUND: FWS AND LAPS

FWS is a federal agency of the U.S. Department of the Interior tasked with protecting and conserving wildlife, plants, and their habitats within the U.S. Along with the National Oceanic and Atmospheric Administration, FWS is the main federal agency responsible for implementing the 1973 Endangered Species Act (ESA).[51] The NWR System Improvement Act of 1997 further expanded FWS's statutory responsibilities to "ensure that the biological integrity, diversity, and environmental health of [ecosystems] are maintained."[52]

FWS's National Wildlife Refuge system protects over 150 million acres of land in 555 refuges and 38 wetland management districts. LAPS has been used by FWS to rank land protection projects for budgeting since 1987. Development of the current version began in 1998 and was used beginning in 2002. FWS currently uses LAPS to prioritize NWRs for receiving funding to buy private lands within their approved acquisition area

---

[50] For recent discussions of portfolio decision analysis, see Salo et al. (2011).
[51] ESA, 16 U.S.C. §1531 et seq. (1973).
[52] See Meretsky et al. (2006).

boundaries. LAPS only ranks refuges that have found willing sellers, and FWS will only approve acquisitions at a "fair market price." According to FWS,

> [LAPS] is based on Service responsibilities and objectives for our Trust resources and legislated responsibilities. Biological priorities are generated using a compilation of 850 possible points, which have been assigned to a comprehensive series of questions. The questions and points assigned to the questions were developed to qualify, quantify, and prioritize Service land protection projects for budget development purposes.[53]

As mentioned above, FWS official responsibilities are to protect and conserve endangered and threatened species, and manage wildlife and ecosystems for the benefit of current and future generations.

Here "Trust resources" refers to trust species, defined by U.S. law as "migratory birds, threatened species, endangered species, interjurisdictional fish, marine mammals, and other species of concern."[54] This motley crew of biological systems (indeed, "species of concern" leaves its extension rather open-ended) reflects the fact (discussed in chapter 1) that identifying biological units worth conservation effort is primarily a normative matter determined by cultural values and political discussion (Norton 1994, Sarkar 2008).

LAPS is a ranking system assigning points to NWRs based on multiple criteria. Of the 850 possible points, there are five criteria, a project summary criterion worth 50 points and four 200-point criteria: Fisheries and Aquatic Resources, Endangered and Threatened Species, Bird Conservation, and Ecosystem Conservation. LAPS thus provides a way for FWS to construct a common scale (construct commensurability) to rank NWRs. Within each criterion are many sub-criteria related to the status of various species and ecosystems as well as the progress of land acquisition on the landscape. Almost all information pertains to features of the landscape located within the NWR's

---

[53] FWS (2011).

[54] 16 U.S.C. §3772 (1).

approved acquisition area, with two exceptions. Contiguous sites with national designations help an NWR's score when those sites contribute to the biological conservation goals of that NWR (e.g. congressionally designated or proposed Wilderness, Ramsar sites, Western Hemisphere Shorebird Reserve Network sites, National Seashores, National Parks, National Monuments, Biosphere Reserves, etc.). Non-designated contiguous protected lands can also add to an NWR's LAPS score.

### 3.2.2. METHODOLOGY

The structure of the descriptive analysis of the next section is as follows. For each criterion, I first describe the goals and values that FWS explicitly claims are associated with that criterion.[55] I then present descriptions of each opportunity for an NWR to score LAPS points. An initial qualitative analysis categorizes points in terms of biocultural, ecological, and management categories. For sub-criteria referring to particular species or ecological types and their properties, I give the classification at the scale of organization given by the LAPS forms. The term 'biocultural' is used here, because units of conservation concern are categorized according to membership in broad taxa (e.g. fish or birds) or non-taxa (e.g. "aquatic species"), along with cultural/legal classifications like "trust species," as opposed to fine-grained strategies of biological taxonomy (e.g. cladistics). 'Ecological' categories include wetland type and habitat type, where the focus is not on single species. Some sub-criteria are non-biocultural and non-ecological, related to management issues like project status, uncertainties, and the like.

There is no single way to place points in categories. For example, should wetland types of concern be distinguished from wetland habitat in general? Should management

[55] Recent LAPS documents (fiscal year 2008) were provided by Deborah Holle, manager of the Balcones Canyonlands NWR. According to LAPS team leader Andrew French (personal communication December 8, 2011), the system has not been changed since then, so the documents are up to date.

actions to improve habitat connectivity in general be lumped together, whether for terrestrial or river systems? I have tried to be faithful to the scale of description given by FWS in the LAPS documents. Conservation of units in these categories do not represent fundamental objectives in the sense given in chapter 2, but rather sub-objectives to FWS's fundamental institutional responsibilities and objectives, which are to conserve wildlife and habitat in general, focusing on "Trust species" or "species of concern" (i.e. species with special cultural, economic, or ecological significance).

A table with the sub-criteria and their associated point values for biocultural, ecological, and management categories, is provided for each criterion as a summary. Two ratios ($R_1$ and $R_2$) are computed for each sub-criterion as a measure of the relative importance of each sub-criterion to its criterion ($R_1$ = total possible sub-criterion points divided by total possible criterion points) and to the overall LAPS score ($R_2$= total possible sub-criterion points divided by 850 total possible points).

While chapter 2 focused on interval scale measurement of value functions, the point system of LAPS allows meaningful ratios to be computed and compared since there are only 850 possible points, and each criterion and sub-criterion has a maximum number of associated points. These ratios can be interpreted as weights for each sub-criterion, representing their relative importance to the final criterion score and final LAPS score, while point values for particular attributes can be interpreted as outputs of value functions. Since LAPS is additive, its implicit functional form is, as in the MAVT models discussed in the last chapter, a weighted sum of the value of each alternative $a_i$ on each criterion $c_j$: $v(a_i) = \sum_j w_j v_j(a_i)$.

For each criterion, I construct an implicit objectives hierarchy for that criterion. Terminal nodes of the hierarchy represent *measurable attributes*, in this case, opportunities to score points. These are associated with sub-objectives higher in the hierarchy, whose achievement helps achieve FWS's more fundamental objectives. (Objectives appear grey in the figures below, while attributes are white.) The complete objectives hierarchy for LAPS (not constructed here due to its size) would connect the objectives hierarchies associated with each criterion to a single fundamental objective for LAPS, which according to FWS is to protect and conserve endangered and threatened species, and manage wildlife and ecosystems for the benefit of current and future generations.

A discussion of each criterion and its associated objectives hierarchy is included before moving to more general normative criticisms in section 4. In particular, I assess the opportunities for an NWR to score LAPS points by the following five desiderata of measurable attributes given in Keeney and Gregory (2005, 3):

1. Attributes should be *unambiguous*: "A clear relationship exists between consequences and descriptions of consequences using the attribute."[56] Attributes that are vague or imprecise are problematic according to this desideratum. Attributes should clearly describe the relevant consequences.
2. Attributes should be *comprehensive*: "The attribute levels cover the range of possible consequences for the corresponding objective, and value judgments implicit in the attribute are reasonable." Comprehensiveness actually contains two desiderata. The first is that attribute levels cover the range of possible consequences, which I will call *comprehensiveness*. The second is that value judgments implicit in attribute levels should be appropriate for the decision problem, which I will call *appropriateness*.

[56] All quotations from Keeney and Gregory (2005, 3).

3. Attributes should be *direct*: "The attribute levels directly describe the consequences of interest." If an attribute describes a consequence that is only partially related to the consequence of interest, it is not sufficiently direct.
4. Attributes should be *operational*: "In practice, information to describe consequences can be obtained and value tradeoffs can reasonably be made." This desideratum states that the availability of data constrains possible attributes. If data is not available to describe a particular consequence, it should not be required by an attribute.
5. Attributes should be *understandable*: "Consequences and value tradeoffs made using the attribute can be readily understood and clearly communicated." In the interest of clear communication, attributes should be understandable by the decision analysts, decision-makers, and interested stakeholders.

Keeney (1992) also distinguishes between natural, constructed, and proxy attributes. Natural attributes are physically measurable or countable quantities with a natural interpretation, for example bird species count or dollars spent. Proxy attributes are like natural attributes (they can be counted or physically measured), however they do not directly measure the achievement of associated objectives, but rather are sufficiently correlated with the objective's achievement to serve as indicators. Thus there may be some uncertainty as to whether a particular proxy attribute is a good indicator of its associated objective, or multiple objectives.[57] One example would be using the presence

[57] This kind of situation may motivate to a two-step multi-attribute analysis that takes into account both a decision-maker's *beliefs* as to the ability of the attributes to indicate the achievement (or non-achievement) of objectives, and then their preferences about how to weight the achievement of such objectives. Butler et al. (2006) discuss this problem and use simulation to evaluate whether an error-prone two-step analysis is worse than an error-prone direct weighting of attributes, finding that the former led to better quality decision even when standard errors for the two-step analysis are twice as high.

of specific vegetation types known to reliably co-occur with a species of conservation concern as an indicator that that species is present. Proxy attributes may fail the desideratum of directness if they are not sufficiently correlated with the objective's achievement. Constructed attributes may be either indices of natural or proxy attributes (like the indices of biodiversity that combine multiple measures discussed in chapter 1), or a constructed scale that measures in a step-wise fashion the achievement of an objective. For example, an ecosystem type in an NWR may be "threatened," "endangered," or "critically endangered" depending on its conservation status, where this categorization may be used as an attribute. These distinctions will be helpful below, particularly in considering questions of means and ends: some attributes in LAPS could be construed as proxy attributes for one objective, or natural attributes for another objective, depending on how the hierarchy is constructed. Where relevant, these issues will be noted below.

As mentioned above, points scored by an NWR are added to produce their final LAPS score. Thus along with an evaluation of the attributes, criterion-specific discussion sections also include assessments of additive independence where appropriate.[58]

[58] Recall from chapter 2 the conditions for additive independence: (i) for all criteria of evaluation $j$ and $k$ where $j \neq k$ and for all alternatives $a_i$, $v_j(a_i)$ does not depend on $v_k(a_i)$; and (ii) for all criteria of evaluation $j$ and $k$ where $j \neq k$ and for all alternatives $a_i$ and $a_m$ where $i \neq m$, $v_j(a_i)$ - $v_j(a_m)$ does not depend on $v_k(a_i)$ or $v_k(a_m)$.

## 3.3. Analysis of LAPS Criteria and Attributes

### 3.3.1. FISHERIES AND AQUATIC RESOURCES

#### 3.3.1.1. Fisheries and Aquatic Resources Declared Values and Objectives

The Fisheries and Aquatic Resources LAPS documents state the FWS's commitment to aquatic habitat and aquatic species conservation and management as legislated responsibilities.[59] Priority is laid on "indigenous or native species within their original ranges and habitats," and aquatic resources and habitats that have been reduced or degraded to "suboptimal levels." The sub-criteria below are all related to either protecting particular populations of fish or other aquatic species, or protecting aquatic habitat in general.

#### 3.3.1.2. Fisheries and Aquatic Resources Sub-criteria

Factor A: Aquatic Resources Population Information (100 points)

A1. Population Status (50 points): Populations of aquatic trust resources in the FWS Fisheries program located within the NWR's approved acquisition area are listed. For each trust species, the manager specifies if the species trend is unknown (0 points), sustainable (5), depleted and candidate (10)—i.e. candidate for listing under ESA, or proposed (15)—i.e. already proposed listing under ESA, for a maximum of 50 points.

A2. Percentage of aquatic Trust species populations benefited by project (50 points): A ratio of the number of species populations listed in A1 to the total number of

---

[59] FWS also discuss section 304 of the Emergency Wetlands Resources Act of 1986, which authorizes the Secretary of the Interior to purchase wetlands of high conservation priority not covered by other legislation, e.g. the ESA or the Migratory Bird Conservation Act of 1929. The law requires the Secretary to consider the "estimated proportion remaining" of wetland types compared to "the time of European settlement," rate of loss and the threat of future losses of various wetland types, and contributions of these types to wildlife, fisheries, water quality, and recreation.

trust species populations on the list for the hydrologic unit[60] in which the NWR is located, is calculated. This ratio is multiplied by 50 to arrive at the score for this sub-criterion.

Factor B: Habitat (100 points)

B1. Life cycles (12 points): Up to four trust species (fish) are listed, with 3 points awarded for life cycle events (nursery, spawning, migration, or wintering) taking place in the NWR.

B2. Barrier removal or passage installation (8 points): If land acquisition will result in the removal of a barrier or installation of a passage that will improve habitat for nursery, spawning, migration, or wintering events for a trust fish species, 2 points each are awarded for improvements for each life cycle.

B3. Free-flowing rivers (10 points): If the project area fully or partially protects a perennial, free-flowing river or a river longer than 125 miles, it receives 10 points.

B4. Wetland types and trends (40 points): Percentage of decreasing wetland types and percentage of former wetlands that will be restored to wetland of a decreasing type found on the NWR are added and multiplied by 40, while percentage of stable wetland types are added and multiplied by 20.[61] (The percentages are converted to ratios between 0 and 1.) These two figures are added and rounded up to the nearest integer, for a maximum of 40 points.

---

[60] Hydrologic units correspond to the boundaries of water drainage systems identified by the United States Geological Survey. Hydrologic units are identified at several scales of classification. The level of classification (sub-region) used by FWS divides 21 regions (the largest classification based on drainage areas of major rivers) into 221 sub-regions. Sub-regions include areas drained by a river system, a closed drainage basin, the extent of a river and its tributaries, or a group of streams in a coastal drainage area.

[61] Decreasing wetland types include varieties of inland, non-tidal wetlands (palustrine emergent, palustrine forested, palustrine scrub-shrub), some varieties of estuary (estuarine intertidal emergent, estuarine intertidal forested, estuarine intertidal scrub-shrub) and marine intertidal zones. Stable wetland types include non-vegetated intertidal estuaries, subtidal estuaries, intertidal, non-vegetated estuaries, sub-tidal estuaries, and lake wetlands.

B5. Wetland losses by state (30 points): A table is included of percent of wetland losses by U.S. state. If the NWR does not include wetlands, it receives no points. If it does, this number is divided by 10, multiplied by 3.25, and rounded to the nearest integer, for a maximum of 30 points.

Table 3.1. Fisheries and Aquatic Resources Summary

| *Sub-criterion* | *Points* | $R_1$ | $R_2$ | *Biocultural categories* | *Ecological categories* | *Management categories* |
|---|---|---|---|---|---|---|
| A1. Aquatic resources population information | 50 | .25 | .06 | Fish. Population trend. Trust. | Aquatic habitats. | Population trend and uncertainty. |
| A2. Percentage of aquatic trust species | 50 | .25 | .06 | Fish. Trust. | Aquatic habitats. | |
| B1. Habitat: Life cycles | 12 | .06 | .014 | Fish. Life cycle events. Trust. | Aquatic habitats. | |
| B2. Habitat: Barrier removal or passage installation. | 8 | .04 | .009 | Fish. Life cycle events. | Aquatic habitats. | Management action to improve habitat. |
| B3. Habitat: Free flowing river | 10 | .05 | .012 | All taxa in rivers. | River habitat. | |
| B4. Habitat: Wetland types and trends | 40 | .2 | .047 | All taxa in wetlands. | Wetland types. Ecological trend (decreasing, stable). | Management action to restore wetlands. |
| B5. Habitat: Wetland loss by state | 30 | .15 | .035 | All taxa in declining wetlands. | Wetlands. | |

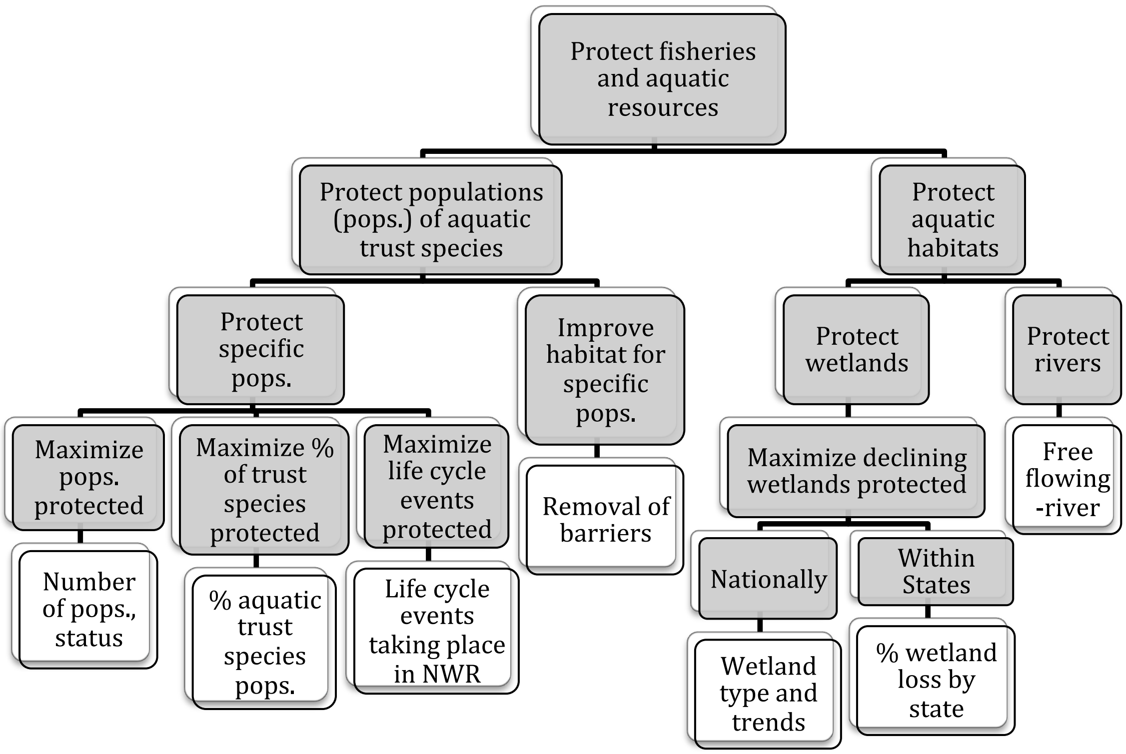



Figure 3.1. Implicit Objectives Hierarchy for Fisheries and Aquatic Resources

### 3.3.1.3. Discussion: Fisheries and Aquatic Resources

The Fisheries and Aquatic Resources criterion attempts to take into account the value of specific populations of aquatic trust species as well as aquatic habitats in general, and wetlands and rivers in particular. Thus two main sub-objectives, “protect populations of aquatic species” and “protect aquatic habitats,” were placed in the implicit objectives hierarchy of Figure 1. These each contained several more sub-objectives each, associated with measurable attributes. Several problems with these attributes are enumerated here, appealing to the desiderata for attributes given above as well as the definition of additive independence given in chapter 2:

1. Attribute A1 specifies that populations of aquatic trust species whose status is unknown are worth 0 points, whereas populations whose status is known to be sustainable are worth 5 points, and more threatened and depleted populations are worth more points. Arguably this attribute is not *appropriate* in that the value judgment that a trust species whose population status is unknown is worth less than a sustainable population (indeed, not worth anything) in the context of this analysis is questionable. We know that such populations count positively for attribute A2; indeed, for attribute A2 all populations of trust species are counted equally in computing the ratio representing local proportional representation of aquatic trust species. While it is reasonable to assess population status in determining the value of a trust species (with more vulnerable populations getting priority), it is unclear why a species whose status is unknown would be worthless here. This is especially puzzling given that in the Endangered and Threatened Species criterion, endangered and threatened species whose population status is unknown are worth more points than those with stable populations; see below.
2. Attribute B3 fails the desiderata of *comprehensiveness*, for the reason that no significant range of values of the objective (protect river habitats) is covered by the attribute, since partial *or* full protection of a perennial, free-flowing river is given the same number of points. The same number of points is given to protection of any river over 125 miles

long. A more comprehensive attribute, which would also be operational, since data would be readily available, would quantify points based on the *quantity* of river habitat protected by the NWR.

3. Recall from chapter 2 the conditions for additive independence, informally that for all alternatives and all criteria, the decision-maker's evaluation of an alternative on a criterion (and the difference between two alternatives on a criterion) does not depend on the value of that (or those) alternative(s) on any other criterion. This is potentially problematic for the separation of aquatic habitat values (e.g. the value of intact rivers or wetlands) and aquatic trust species values, since the biological conservation value of an NWR for its habitat presumably depends on the occurrence of various species in that habitat. (The issue of additive independence is relevant because the point values of these sub-criteria are added to produce an overall criterion score.) While it may be valuable to maintain viable habitat in the absence of trust species occurrence, the value of the former at one level is presumably higher with species occurrence. Additive independence may still hold within the range of habitat and species values relevant to this analysis, however. Assessment of FWS managers' judgments within the range of values for habitat and species sub-criteria would have to be performed to verify local additive independence.

4. Finally, in the hierarchy above attribute B1 related to life cycles was attached to the objective "maximize number of life cycle events protected," on the assumption that this sub-objective helps achieve the more general sub-objective of protecting specific populations of aquatic

trust species. Construed this way, B1 is a natural attribute for this sub-objective. However it could also be taken as a proxy attribute for an objective that is less *directly* related, for example protecting aquatic habitat or protecting specific populations, on the assumption that decision-makers do not care directly about maximizing life cycle events that take place in NWRs. However, such concern is not unreasonable in this context, since the more life cycle events that take place in a protected area, the more overall protection is afforded to the relevant populations.

### 3.3.2. Endangered and Threatened Species

#### 3.3.2.1. Endangered and Threatened Species Declared Values and Objectives

The Endangered and Threatened Species LAPS documents state the FWS's goals (1) "to conserve and recover listed, and proposed species"; (2) "to protect habitats on which listed and proposed species depend"; and (3) "to restore depleted populations and/or habitat to preclude the necessity for listing actions." As mentioned above, FWS is the main federal agency charged with implementing ESA.

#### 3.3.2.2. Endangered and Threatened Species Sub-criteria

Extinction prevention/de-listing (200 points): If an NWR will prevent the extinction of any listed species, or would effectively recover a listed species such that it may be de-listed, that NWR is rewarded the maximum 200 points for this category. These claims require extensive documentation and support from managers in charge of implementing recovery plans.

For factors A-C, managers give all listed species in the NWR that do not qualify for extinction prevention/de-listing status.

Factor A: Population Information (16 points maximum per species)

A1. Listing status (2 points): Endangered or proposed endangered species are given 2 points, threatened or proposed threatened species get 1 point.

A2. Population trend (4 points): Species with declining populations get 4 points, if the trend is unknown it is given 3 points, for stable trend 2 points, and increasing 0.

A3. Recovery achieved (4 points): Percentage of species recovery objectives achieved determines points: 0-25%: 1 point; 26-50%: 2; 51-75%: 3; 76-100%: 4.

A4. Recovery priority number (6 points): Since 1983, the FWS has had a recovery priority system to rank listed species. The rank determines points: 1-3: 6 points; 4-6: 4; 7-12: 2; 13-18: 1.

Factor B: Habitat Use Description (5 points per species)

B1. Habitat use: If a species is resident, it is given 5 points. If seasonal, 3 points. If occasional, 0 points.

Factor C: Rationale (11 points per species)

C1. Population goal (2 points): If an NWR maintains a project to maintain the population, it is given 0 points. If the project serves to increase population, it is given 2 points.

C2. Completion of recovery plan objectives (5 points): 5 points for completion of at least one objective in recovery plan.

C3. Habitat restoration need (4 points): None: 4 points; Low: 3; Moderate: 2; High: 0.

Factor D: Additional Information

D1. If the project is part of another listed species' historical range suitable for reintroduction, 2 points per species are awarded.

D2. If the project addresses needs of species identified in the Candidate Notice of Review (candidates to be listed), 5 points per species are awarded.

Table 3.2. Endangered and Threatened Species Summary ($x$ = number of endangered and threatened species; $y$ = number of other listed species targeted for reintroduction; $z$ = number of candidates for listing.)

| *Sub-criterion* | *Points* | $R_1$ | $R_2$ | *Biocultural categories* | *Ecological categories* | *Management categories* |
|---|---|---|---|---|---|---|
| Extinction prevention/de-listing | 200 | 1 | .235 | Listed (endangered, threatened) species. | | Prevention of extinction. |
| A1. Listing Status | $2x$ | $.01x$ | $.002x$ | Listed species. | | |
| A2. Population trend | $4x$ | $.02x$ | $.0047x$ | Population trend (declining, stable, improving). Listed. | | Population trend and uncertainty. |
| A3. Recovery achieved | $4x$ | $.02x$ | $.0047x$ | Listed. | | Management goals achieved. |
| A4. Recovery priority | $6x$ | $.03x$ | $.007x$ | Listed. | | Recovery priority. |
| B1. Habitat use | $5x$ | $.025x$ | $.0059x$ | Listed. Seasonality of habitat use. | | |
| C1. Population goal | $2x$ | $.01x$ | $.002x$ | Listed. | | Management goal. |
| C2. Plan objectives completion | $5x$ | $.025x$ | $.0059x$ | Listed. | | Management goals achieved. |
| C3. Habitat restoration need | $4x$ | $.02x$ | $.0047x$ | Listed. | Habitat for listed species. | |
| D1. Other listed species historic range | $2y$ | $.01y$ | $.0047y$ | Listed. | Habitat for listed species. | |
| D2. Candidate notice species | $5z$ | $.025z$ | $.0059z$ | Candidates for listed species. | | |

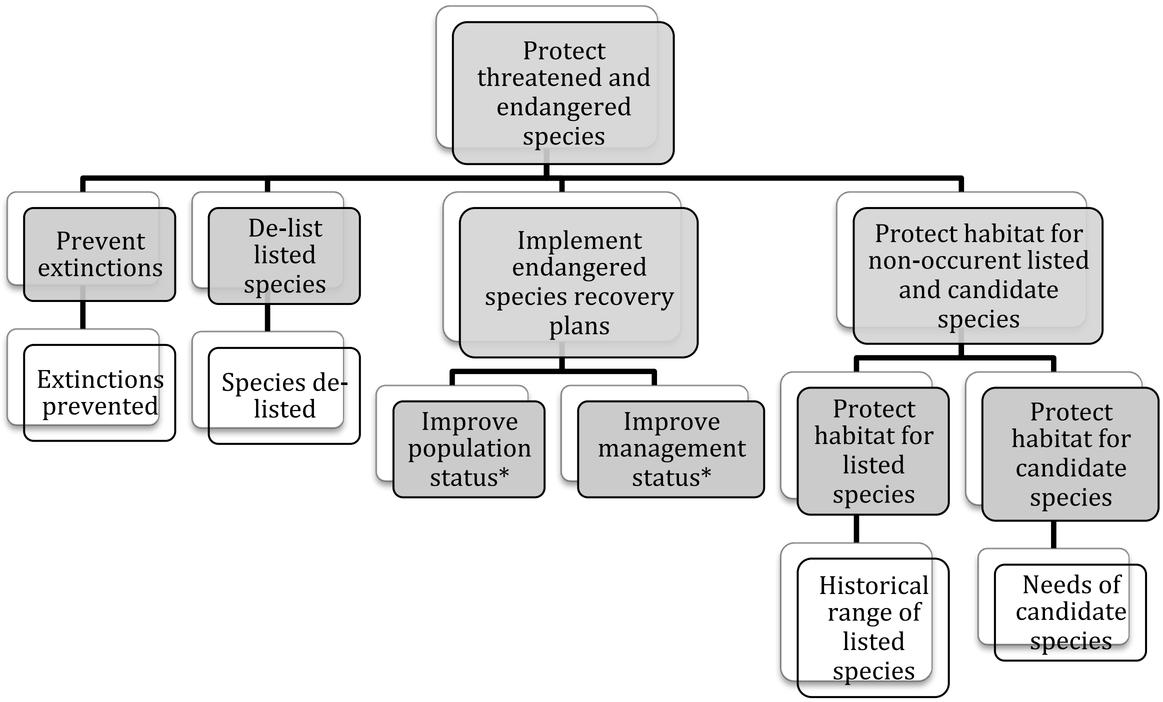



Figure 3.2.a. Implicit Objectives Hierarchy for Endangered and Threatened Species.

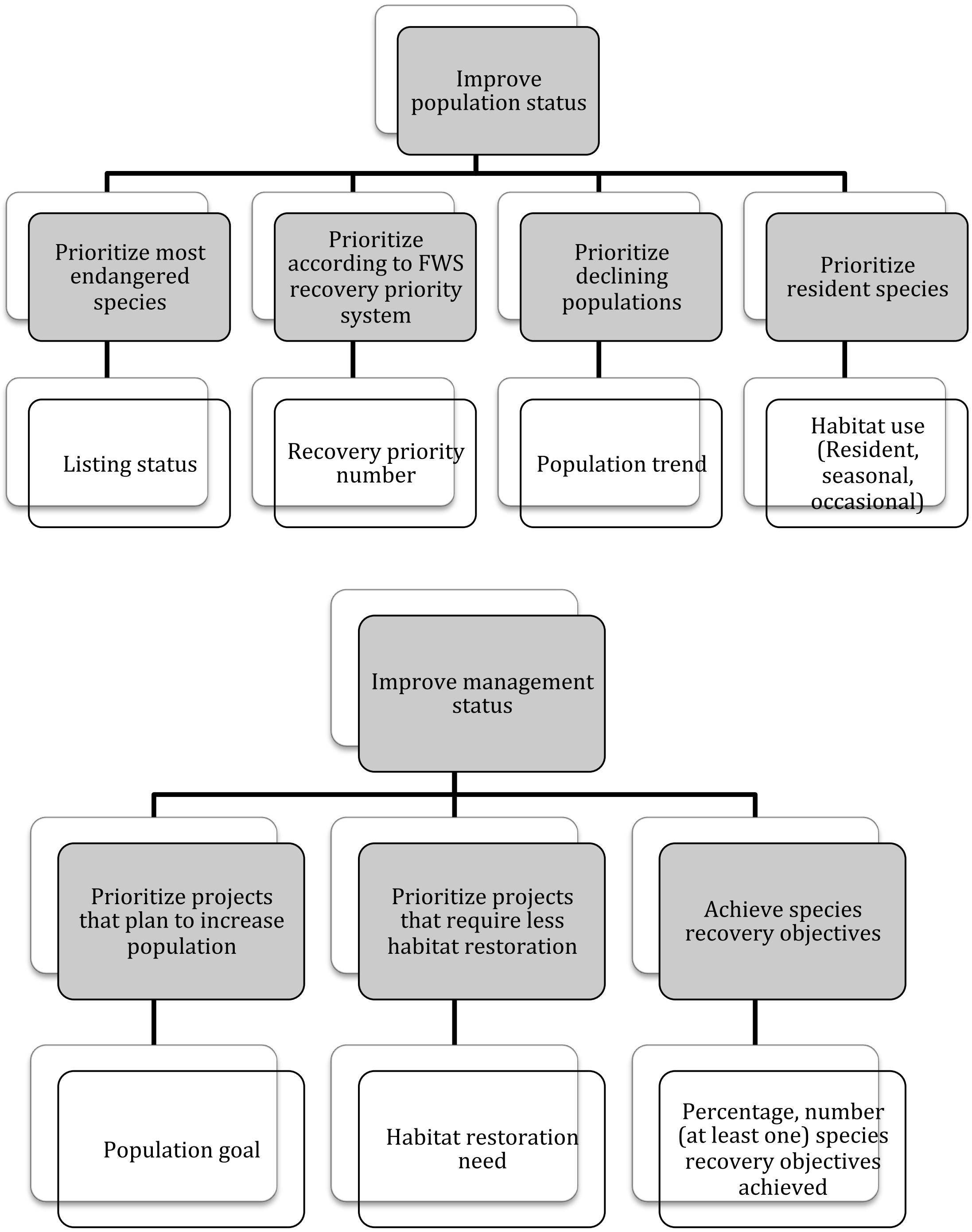



Figure 3.2.b. Sub-objectives and Attributes Related to Improve Population Status and Improve Management Status.

#### 3.3.2.3. Discussion: Endangered and Threatened Species

The Endangered and Threatened Species criterion takes into account an NWR's endangered and threatened species, especially whether an NWR will prevent extinction of an endangered species or will lead to a species being de-listed, since NWRs preventing extinction or leading to a de-listing earn all 200 possible points for this criterion. Potential problems with the implicit objectives hierarchy for this criterion include the following:

1. The attributes for the prevention of extinction and de-listing represent failures of *comprehensiveness* and *appropriateness*. Firstly, the range of values at stake is not represented by the range of values of the attributes, since only one extinction prevention or de-listing can earn the maximum LAPS points for this criterion, whereas it is conceivable that de-listing *and* extinction prevention, or multiple de-listings, or multiple extinction preventions could occur at a single NWR. As the system stands, these more favorable outcomes would only be eligible for the same number of points. Secondly, the value judgment implied by the same number of points being given to extinction prevention and a de-listing is questionable in this context. Other attributes of LAPS prioritize species that are more rare (e.g. in the fisheries and aquatic resources sub-criteria). Even within this criterion, A1 gives more points to endangered than threatened species, and A2 gives more points to species whose populations are declining. This would seem to imply that de-listings should be given fewer LAPS points than the prevention of extinctions.
2. The Population Information and Rationale sub-criteria (represented in the objectives hierarchy as "improve population status" and "improve management status") contain redundant attributes, since A3 and C2 both

measure the achievement of recovery plan objectives. A3 is more comprehensive, since it covers a wider variation in the achievement of recovery plan objectives, while C2 only awards points when at least one objective is achieved. Redundancies lead to failures of additive independence, since the score on one attribute is not independent of the score on another. The system could also be made simpler and more understandable if redundancies were removed.

3. Other potential failures of additive independence should be noted, particularly between attributes under population status. Listing status, population trend, and recovery number are assumed to be additive independent here. But the value of a threatened or endangered species could depend non-additively on its population trend and/or how its recovery has been prioritized by FWS. For example, population trend points are the same for threatened or endangered species. Thus an endangered species that is declining gets the same points for that decline as a proposed threatened species. Again, additive independence may still hold for these particular ranges of values.

### 3.3.3. Bird Conservation

#### 3.3.3.1. Bird Conservation Declared Values and Objectives

The Bird Conservation LAPS documents state that the Bird Conservation category points are intended to measure the importance of an NWR to bird species of conservation concern and to avian diversity in general. Lists of bird species of conservation concern at the national and regional levels ("Bird Conservation Regions" or BCRs) are compiled every few years by FWS. This component of LAPS is supposed to determine "the importance of an NWR to populations, species, and diversity" of birds at these regional and national levels. Listed birds are not included in this section, as they are covered by the Endangered and Threatened Species criterion.

**3.3.3.2. Bird Conservation Sub-criteria**

Factor A. Importance to Specific Populations (100 points)

A1. 100 points if an NWR supports at least 50% or more of the overall population for any bird species in North America (except listed species) for at least one life cycle period.

A2, A3. For an NWR that supports 5-49% of the overall population of a bird species in North America (except listed species) for at least one life cycle period, if that species is recognized by FWS as a bird species of conservation concern for that BCR, 40 points, if not, 20 points.

Factor B. Importance to Priority Species (80 points)

B1. A "species importance value" is calculated by dividing the number of species on the regional BCR list that use the NWR as habitat for at least one life cycle period by the total number of species on the list, and multiplying this ratio by 80.

Factor C. Avian Diversity (20 points)

C1. An avian diversity score is calculated by dividing the number of bird species on the national BCR list that use the NWR as habitat for at least one life cycle period by the total number of species on that list, and multiplying this ratio by 60, for a maximum of 20 points.

Table 3.3. Bird Conservation Summary

| *Sub-criterion* | *Points* | $R_1$ | $R_2$ | *Biocultural categories* | *Ecological categories* | *Management categories* |
|---|---|---|---|---|---|---|
| A1. Importance to specific | 100 | .5 | .118 | All bird species. | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| populations (50-100%) | | | | | | |
| A2. Importance to specific populations (5-49%): regional concern | 40 | .2 | .047 | Birds on regional BCR list. | | |
| A3. Importance to specific populations (5-49%): not regional concern | 20 | .1 | .024 | All bird species. | | |
| B1. Importance to priority species | 80 | .4 | .094 | Birds on regional BCR list. | | |
| C1. Avian diversity score | 20 | .1 | .024 | Birds on national BCR list. | | |

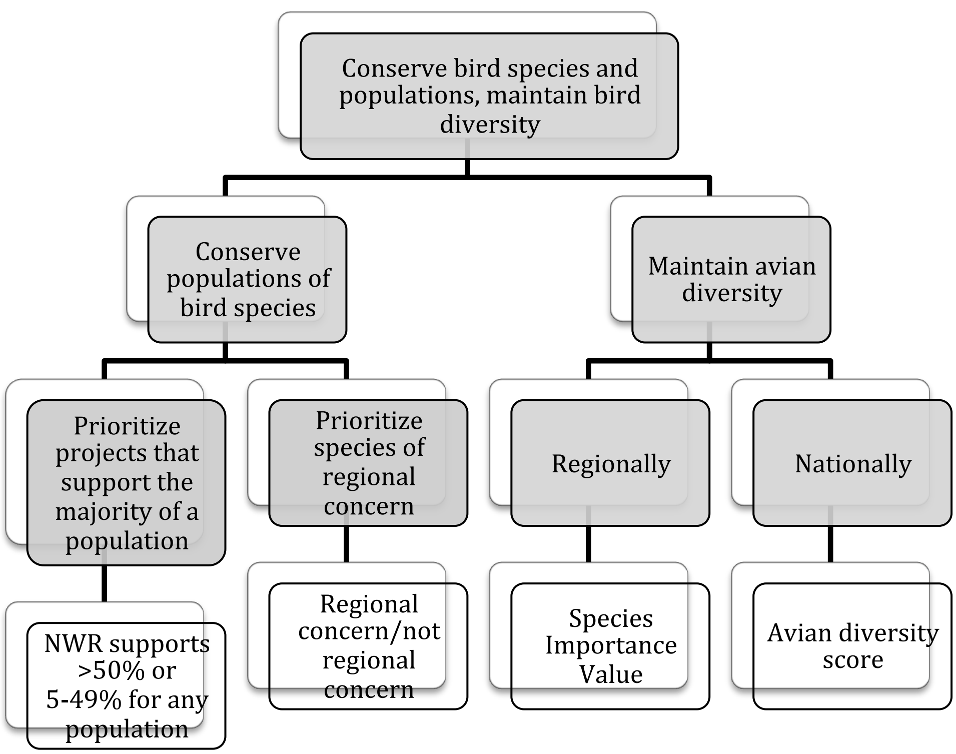



Figure 3.3. Implicit Objectives Hierarchy for Bird Conservation

### 3.3.3.3. Discussion: Bird Conservation

The Bird Conservation criterion takes into account an NWR's bird populations, as well as proportional representation or diversity of birds of conservation concern. It is a much simpler criterion than the Fisheries and Aquatic Resources criterion since it does not also attempt to take habitat into account, and there are no management objectives or considerations of uncertainty. However, the following problems should be noted:

1. The "species importance value" and "avian diversity score" attributes depend on the same type of ratio being computed, namely a ratio between the number of birds on the national or regional BCR list, respectively, that use the NWR for at least one life cycle and the total number of birds on the regional or national BCR list. Since both of these ratios describe avian diversity at different scales, they were placed under the sub-objective "Maintain avian diversity" in the objectives hierarchy above. In order to improve comprehensibility, these attributes should be renamed in a uniform way, for example "regional avian diversity score" and "national avian diversity score." Species on the national list also appear on regional lists, so those particular species are counted twice for NWRs where they occur. The rationale for this double counting is that these ratios take into account the NWR's contribution to regional and national avian diversity, which may be different. (Contribution to regional avian diversity is weighted more than contribution to national avian diversity: 80 versus 20 possible points.)
2. Here and elsewhere in the system (e.g. in wetland type and trends in the Fisheries and Aquatic Resources criterion), ratios are multiplied by a certain number of points in order to arrive at a final score for that attribute. Here, the maximum points possible for each attribute add up to more than 200, so the

maximum number of points achievable for the ratio-computed attributes must be lower than the numbers multiplied by the ratios. (In the case of regional and national avian diversity, these numbers are 80 and 60.) This makes the range of values one would expect for these attributes obscure. A more comprehensible attribute would display the range of possible ratios and associate each with appropriate point values.

### 3.3.4. ECOSYSTEM CONSERVATION

#### 3.3.4.1. Ecosystem Conservation Declared Values and Objectives

The Ecosystem Conservation LAPS documents state that the category points are intended to measure "opportunities to effectively conserve and protect endangered and threatened ecosystems, and large, intact habitats, as a means of promoting and perpetuating the Service's trust resources." Thus FWS Trust resources (species) are conceived as the relevant value and ecosystem conservation as a means.

#### 3.3.4.2. Ecosystem Conservation Sub-criteria

Factor A – Landscape (60 points)

A1. Size of landscape effort (55 points): This may include land owned by private partners, for example private landowners with conservation easements, or other public agencies. Conservation projects up to 50,000 acres get 0 points. 50,000-125,000 acres: 5 points; 125,000-200,000: 10; 200,001-275,000: 15; 275,001-350,000: 20; 350,001-425,000: 25; 425,001-500,000: 30; 500,001-575,000: 35; 575,001-650,000: 40; 650,001-725,000: 45; 725,001-800,000: 50; 800,001- : 55.

A2. Partnerships: Land protection partnerships earn 5 points. These are quantifiable commitments to land protection made by partners (e.g. states or non-governmental organizations).

Factor B: Ecosystem (75 points)

For this factor, published work on the conservation ecology of the United States (for example, Noss et al. 1995 or the works cited therein) may be used to cite critically endangered (greater than 98% decline, 25 points each), endangered (85-98% decline, 20 points each), and threatened (70-84% decline, 15 points each) ecosystems within an NWR's approved acquisition area, for a maximum of 75 points. For example, if an NWR contained three critically endangered ecosystems, it would score the maximum 75 points, whereas if it contained one endangered and one threatened ecosystem type it would score 35 points.

B1. Critically endangered ecosystems: 25 points each.

B2. Endangered ecosystems: 20 points each.

B3. Threatened ecosystems: 15 points each.

Factor C: Site (65 points)

C1. Project size (45 points): The greater the NWR acreage within the whole landscape effort, the more points it scores. 0-9,999 acres: 0 points; 10,000-19,999: 5; 20,000-24,999: 10. Points increase linearly by 5 points per 5,000 acres (e.g. 40,000-44,999 acres is worth 30 points, 45,000-49,999 acres is worth 35 points) until they increase 10 points for projects ≥ 50,000 acres, worth the maximum 45 points.

C2. National designations: If an NWR is part of, or contiguous to a site that has been designated of national importance and that biologically contributes to the objectives of the project, 5 points are awarded for each designation. These include congressionally designated or proposed Wilderness, Ramsar sites, Western Hemisphere Shorebird Reserve Network sites, National Seashores, National Parks, National Monuments, Biosphere Reserves, etc.

Table 3.4. Ecosystem Conservation Sub-criteria Summary (*a* = number of critically endangered ecosystems; *b* = number of endangered ecosystems; *c* = number of threatened ecosystems).

| *Sub-criterion* | *Points* | $R_1$ | $R_2$ | *Biocultural categories* | *Ecological categories* | *Management categories* |
|---|---|---|---|---|---|---|
| A1. Size of landscape effort | 55 | .275 | .065 | | | Size of effort on landscape. Partnerships. |
| A2. Partnerships | 5 | .025 | .006 | | | Partnerships. |
| B. Ecosystem | 75 | .375 | .088 | | Imperiled ecosystems. | |
| B1. Critically endangered ecosystems | 25*a* | .125*a* | .029*a* | | Critically endangered ecosystems. | |
| B2. Endangered ecosystems | 20*b* | .1*b* | .024*b* | | Endangered ecosystems. | |
| B3. Threatened ecosystems | 15*c* | .075*c* | .018*c* | | Threatened ecosystems. | |
| C1. Project size | 45 | .225 | .05 | | | Size of NWR. |
| C2. Contributions to national designations | 20 | .1 | .023 | | | National designations. |

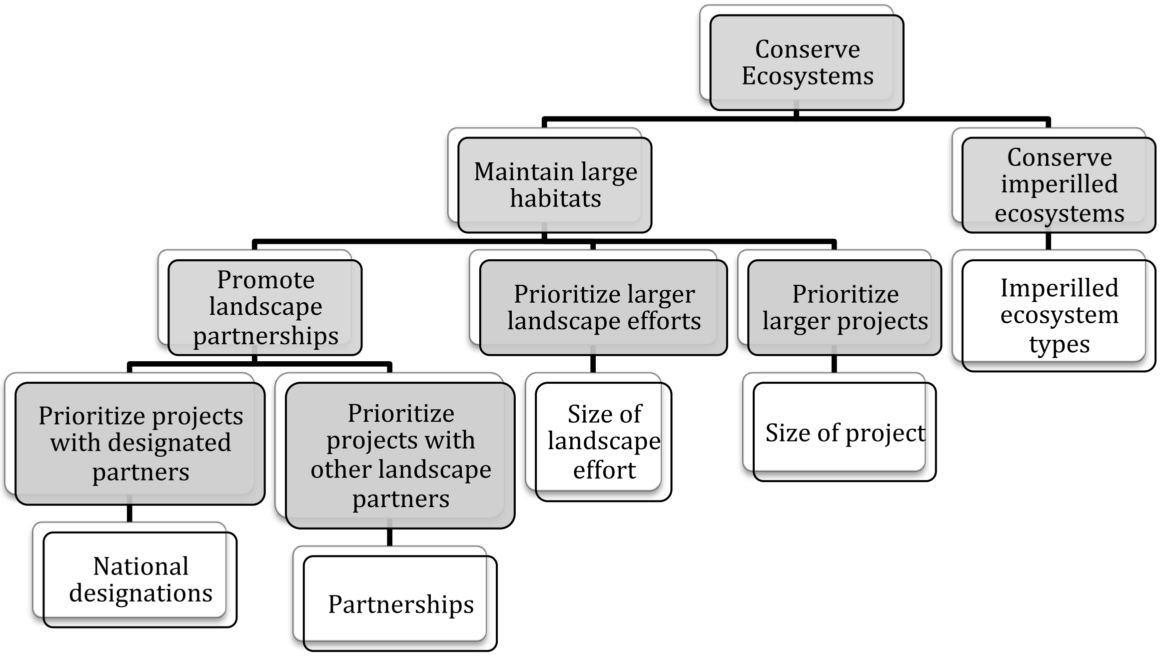



Figure 3.4. Implicit Objectives Hierarchy for Ecosystem Conservation

#### 3.3.4.3. Discussion: Ecosystem Conservation

The Ecosystem Conservation criterion takes into account imperiled (threatened, endangered, or critically endangered) ecosystems contained in an NWR, the total size of the landscape effort and the size of the NWR, and partnerships across the landscape. Issues with this criterion include the following:

1. One potential failure of additive independence should be noted, namely that FWS's valuation of the size of the landscape effort and size of the project may depend non-additively on the occurrence of imperiled ecosystems, or the occurrence of species of conservation concern from other criteria (larger species generally require larger areas of contiguous habitat). A very large NWR that does not contain any imperiled ecosystems will receive the same number of points for its size as an equally sized NWR with many imperiled

ecosystems. Again, this is only a potential failure since additive independence may still hold within the range of values taken by these attributes.

2. The attribute for imperiled ecosystem types fails of *comprehensiveness*, since the point values do not take into account the range of consequences, namely quantities of the imperiled ecosystems. An NWR with a large intact tract of critically endangered habitat will receive the same number of points as an NWR with a relatively smaller intact tract of critically endangered habitat. This information may not be contained in the general information about the project size and the size of the total landscape effort.
3. One last minor issue related to the structure of the objectives hierarchy should be mentioned. The attributes related to national designations and partnerships were placed under the objective “promote landscape partnerships,” which in turn contributes to the objective “maintain large habitats.” While it is true that landscape partnerships contribute to large habitats, they may have value independent of this contribution, since they engage other stakeholders, including federal agencies, conservation organizations, and private landowners. For example, promotion of such ties to other stakeholder groups helps distribute the benefits and burdens of conservation more widely across government and society generally. This issue will be discussed below, where I argue that more detailed social scientific information could usefully be incorporated into LAPS.

### 3.3.5. PROJECT SUMMARY

The Project Summary component contains "bonus" points for the completion of projects. It gives 50 points to NWRs that have acquired more than 95% of their land, 25 points to NWRs that have acquired 90-94.9% of their land, and 10 points for 80-89.9%.

Since this is only a single attribute I do not construct a Project Summary objectives hierarchy.

Table 3.5. Project Summary

| *Sub-criterion* | *Points* | $R_1$ | $R_2$ | *Biological categories* | *Ecological categories* | *Management categories* |
|---|---|---|---|---|---|---|
| Degree of completion | 50 | 1 | .059 | | | Completion of acquisition. |

### 3.3.6. LAPS CRITERIA GENERAL DISCUSSION

#### 3.3.6.1. Additive Independence

Since LAPS points are added across criteria and sub-criteria to produce a final score, additive independence should hold between them along the ranges given in the attributes. If additive independence fails, FWS's preferences cannot be represented by an additive function like the one implied by LAPS. Several potential failures of additive independence were noted above, based on plausible lines of practical reasoning derived from valuations implicit in LAPS. However it should be stressed that without more information about FWS managers' valuations over variations in the attributes, no firm conclusions about additive independence can be made, since additive independence is only required within the given ranges of the relevant attributes.

#### 3.3.6.2. Management and Non-management Attributes

The four main sub-criteria are capped at the maximum of 200 points, meaning that considerations involving fish and aquatic resources, birds, threatened and endangered species, and ecosystems cannot count for more than 23.5% (200/850) of total points each. Within these four categories (800 points), however, management variables related to the size of the project, partnerships, completion of management goals, national designations, etc. (excluding extinction preventions and de-listing actions) account for at least 172 points, or roughly 20% of total points. Including the 50-point project summary, 222 points or 26% of total points are specifically related to these management variables, which are only indirectly related to particular taxa or ecosystems. The only criterion that does not include any management variables is the Bird Conservation criterion, whose sub-criteria are only related to bird species richness and diversity. It would be consistent with the implicit valuations in LAPS if management attributes relevant to bird conservation were included in the Bird Conservation criterion.

#### 3.3.6.3. Variability in Species' Value

In LAPS, there is very high variability in the value of species. As mentioned above, a single bird species the population of which resides almost completely in an NWR (but is not endangered or threatened) can be worth 100 points (11% of total points), whereas a bird that is not of regional conservation concern may be worth 20 points (2% of total points). If an NWR is preventing the extinction of a listed species, or delisting a species, 200 points are awarded (23.5% of total points). However, for endangered or threatened species that are not eligible for de-listing or extinction prevention, each species listed is worth at most 32 points (including related management variables, roughly 4% of total points). It does not seem consistent with other aspects of LAPS that a bird of conservation concern that is not listed could be worth more than three times as many points as a listed species not eligible for delisting or extinction prevention. On the other hand, a fish trust species may be worth as little as 8 points (just below 1% of total points) if its population is sustainable and it completes only one life cycle stage in the NWR. Much of this variability likely reflects legitimate variability in FWS managers' and biologists' valuations of these species. However, some of it might be an artifact of the system's construction. Since there is no method internal to LAPS that justifies individual species' contributions to an NWR's conservation value, one can only speculate.

#### 3.3.6.4. Population Trends

The Fisheries and Aquatic Resources and Endangered and Threatened Species criteria both include information about the trends of populations. More points are awarded for NWRs with species on the decline. For Endangered and Threatened Species, only 4 points for each species relate to population trends. For Fisheries and Aquatic Resources, depleted trust species receive 10 points each. No information about population trends appears in the Birds criterion. If population trend data is available for birds on the BCR lists, it would be consistent with the implicit valuations in LAPS's other criteria to include this information in the Bird Conservation criterion.

## 3.4. Normative Criticisms of LAPS

### 3.4.1. Criteria and Point Values

Inspection of the LAPS documents themselves does not reveal the method by which points were allocated to the different sub-criteria and attributes, and the method by which the sub-criteria were constructed for each criterion is equally unclear. However, according to FWS LAPS team leader Andrew French, the method used to assign point values was consensus between representatives of the nine regional FWS offices, and values were assigned on the basis of importance of the particular criterion to accomplishing the overall objectives and statutory responsibilities of FWS.[62] According to French, at least 25 people were present at these meetings, and sensitivity analysis was performed to make sure that the overall rankings produced by the system made intuitive sense to those present.

However, some arbitrariness in the assignment of points can be seen in the variability in the value of particular species, ranging from 8 points to 100 points for non-threatened or endangered species, and 200 points for species on the brink of extinction. Except for certain obvious rules, for example that rarer species are worth more points than less rare ones or that individual bird species are usually worth more than fish species, it is difficult to discern any patterns of point-assignment within or between categories. Complicated rules for calculating point values (see, for example, the calculation of points for wetland losses by state) exacerbate this problem.

One more note about point values concerns means and ends. In the LAPS documents for the Ecosystem criterion, FWS state that ecosystem conservation is a *means* to conserve trust species. If this is so, then habitat conservation variables should be included in each of the criteria dealing with trust species. However, given the points awarded to rare and declining ecosystem types, independent of the presence of particular trust species, this does not seem to be the case.

### 3.4.2. No Complementarity

Although it is intended to rank NWRs for the acquisition of lands within their approved acquisition area boundaries, LAPS only performs an overall assessment of the NWR. Thus the marginal benefit to biological conservation of a new acquisition, or its

---

[62] Personal phone interviews with Andrew French, December 8, 2011 and January 17, 2012.

*complementarity value* to the network of NWRs (Margules and Sarkar 2007, Sarkar 2012b), is not calculated by the system. This means that an NWR with high overall value that is proposing to purchase degraded land may be prioritized over an NWR with low overall value that is proposing to purchase a parcel of land with much higher biological value. By hypothesis, the marginal benefit to conservation of the latter purchase would be higher than the former, however the system is set up such that the former would be prioritized for acquisition. Furthermore, since "fair market prices" for real estate will differ greatly in different parts of the country, a criterion of cost along with complementarity value is necessary for economical prioritization.

### 3.3.2. Well-roundedness

Because each criterion includes a maximum number of points, the system rewards NWRs that are well-rounded, or do well enough on all categories. However, this precludes the possibility that NWRs may have more focused priorities, for example on ecosystem-type or bird conservation. It is possible that FWS's institutional objectives may best be served by investing in several projects that would have very high values on particular sub-criteria were those scores not capped, but whose overall score is lower than others with those maxima in place. One of the main purposes of constructing a common scale is to identify situations like this, where low rankings on one criterion may be traded off against larger benefits in other criteria. The structure of LAPS precludes this for certain ranges of biological and ecological value. It could be the case that the system was designed to prevent this by making it nearly impossible to score the maximum points. A perusal of the rankings for 2013 Fiscal Year indicate that although this is true for the Bird and Fish criteria, 200 points were scored by 13 NWRs in the top 30 for Endangered and Threatened Species criterion, and 200 points were scored in the Ecosystem criterion by 8

NWRs in the top 30. If those NWRs were allowed to score more than 200 points in those categories, their rankings may have been different.

### 3.4.4. LAPS and Portfolio Selection

The problems highlighted in the previous two paragraphs are both related to the fact that LAPS was not designed to choose one best NWR to fund a single land acquisition project, but rather to produce an overall ranking of NWRs to assist in the prioritization of many possible land acquisition projects. The overall decision problem actually faced by FWS is thus best represented as a "portfolio selection problem" (Salo et al. 2011), where a finite subset of alternatives is chosen from a larger set of alternatives.[63] LAPS does not solve the portfolio selection problem on its own, since it does not calculate the marginal value of acquisitions or provide a way of choosing a subset. However it should be noted that problems arise for the use of MAVT in the portfolio selection problem, in particular for the use of additive models like that in LAPS, due to a failure of additive independence.[64] Evaluating an alternative for inclusion in the portfolio (the preferred subset of projects that will be pursued) will not in general be additive independent of the valuations of other alternatives in the portfolio. In the context of FWS's decision problem, the marginal value of a new acquisition, or its complementarity value to the network of NWRs, will in most cases depend on the value of other acquisitions being pursued. For these reasons, LAPS alone cannot be used to select a

---

[63] Salo et al. (2011) provide background and several applications of decision analysis applied to the portfolio selection problem. For an early application of MAVT to the portfolio selection problem, see Golabi et al. (1981). For a review of quantitative models of project selection (in the context of research and development projects), see Heidenberger and Stummer (1999).

[64] See also Clemen and Smith (2009), who show that, in applying MAVT to the portfolio selection problem, different ways of defining the baseline valuation for the status quo, or *not* pursuing a project, can affect which projects are prioritized in ways that may not reflect the decision-makers' valuations.

portfolio of land acquisition projects, but must be supplemented by other decision support tools.

### 3.4.5. No Social Scientific Data

With the notable exception of the partnerships sub-criterion in the Ecosystem criterion (only worth 5 points or 0.5% of total points), there are no criteria that take social scientific data into account, related to risks (from economic development, pollution, or local hostility) or benefits to surrounding communities (education, outreach, economic impact, or other stakeholder involvement). There are ethical and practical arguments for including latter, positive aspects of stakeholder involvement. The ethical argument is that citizen stakeholders fund FWS for the purpose of serving those stakeholders and future generations. FWS's mission statement explicitly assumes this ethical responsibility. The practical argument refers to the possibility that some NWRs may have lower needs to purchase land if private landowners may be trusted to manage their land in accordance with FWS objectives. Social scientific data on conservation attitudes and knowledge would be useful here.[65] On the other hand, data on threats would also be useful, since the success of an NWR's biological conservation goals often may depend on these factors (e.g. the presence of polluters or rural land development, or local hostility toward the refuge's goals).

## 3.5. Conclusions and Transition

### 3.5.1. LAPS and Constructing Commensurability

The last chapter made an argument that constructing commensurability, or constructing preferences over alternatives in the presence of complex tradeoffs between values, is a requirement of practical rationality for important decisions. The decisions

[65] For an example of this kind of study, see Knight et al. (2010).

LAPS was designed to help FWS deal with are good examples of decision problems where the requirement of constructing commensurability holds. Prioritizing NWRs for land acquisition budgeting must take into account FWS's many institutional responsibilities and objectives that correspond to many different values. The stakes are high, so it makes sense for FWS to spend time and energy systematically prioritizing NWRs. LAPS provides a way of constructing commensurability between, for example, the value of preventing the extinction of an endangered species and the value of maintaining intact wetland habitats.

Most of the numerous value judgments implicit in LAPS's tradeoffs between the values at stake were not examined here, however my discussions above noted areas where LAPS seemed inconsistent in its valuations. A more comprehensive normative assessment of LAPS would thoroughly examine all implicit value judgments, including tradeoffs. Apparent inconsistencies may be due to the fact that FWS managers and biologists devised LAPS without the consultation of decision analysts. Thus while it could be interpreted as a multi-attribute value function, LAPS's system for assigning points was not constructed with full consideration of the constraints on such functions given in chapter 2, in particular additive independence. However, before a multi-attribute value function could be constructed, FWS would need a clear and unambiguous objectives hierarchy including attributes that better satisfy the desiderata enumerated above. The construction of such a revised objectives hierarchy would be a worthwhile project for future research, but would require extensive consultation with FWS managers and biologists.

### 3.5.2. From Decisions with Multiple Values to Decisions with Multiple Agents

One reason given for the use of social scientific data in LAPS was that the actions of local landowners and other agents in communities surrounding NWRs affect conservation outcomes, whether positively or negatively. When multiple agents' actions affect an outcome of a decision, it is not sufficient for rational choice under certainty to consider merely one's own valuations over alternatives. Agents must develop expectations of what other agents will do under the assumption that other agents have their own values and are similarly developing such expectations. This kind of strategic thinking is formalized in the theory of games. While this chapter and the last were concerned with normative applications of multi-attribute decision theory in conservation problems with multiple values, the next chapter will identify a normative role for game theory in conservation problems with multiple agents.

# Chapter 4: Conservation Dilemmas: *Game Theory, Group Decisions, and the Limits of Mechanism Design*[66]

## 4.1. Introduction

### 4.1.1. A Normative Role for Game Theory

Biological conservation efforts require several normative commitments. First, as discussed in chapter 1, operationalizing the concept 'biodiversity' involves choosing taxa and measures of heterogeneity that contribute to our goals and values. Second, the goal of biological conservation must be negotiated with other normatively salient social goals such as economic welfare, public health, etc. In these and other cases, there is much potential for conflict. When these conflicts occur for a single agent (individual or organization), decision support tools based on multi-criteria analysis, like those discussed in chapter 2, can provide useful insight (Moffett and Sarkar 2006). When these conflicts involve multiple groups or agents, game theory can play a parallel role.

In conservation contexts, two potential roles for game theory should be distinguished. The first role, well understood in evolutionary theory and economics, is descriptive. Evolutionary games model frequency-dependent selection; in economics, traditional game theory is used to explain macro-behavioral outcomes by appealing to the equilibrium of some underlying game (Osborne and Rubinstein 1994, Gintis 2009). Game theory can similarly be used to describe conservation conflicts.

However, the focus here will be on a second, *normative* or prescriptive role of game theory: identifying "dilemma" conflicts with Pareto-inefficient Nash equilibria[67]

[66] Sections 4.1, 4.2, 4.3.1, and 4.3.3 of this chapter are based on joint work with Sahotra Sarkar, see Frank and Sarkar (2010).

[67] Throughout I will use the following standard definitions of Nash equilibrium and Pareto-efficiency. An outcome is a *Nash equilibrium* if no agent can do better by *unilaterally deviating* from the current strategy: each agent's action is a "best response" to the actions of the other agents. An outcome is *Pareto-efficient* if, relative to the other possible outcomes, no agent can be made better off without making at least one agent

can enable constructive action in order to achieve (closer to) optimal conservation outcomes, whether by familiar mechanism-design-style policy solutions or otherwise.[68] Indeed, attaining cooperative outcomes need not proceed via formal institutional arrangements at all, but may be achieved through group deliberation and the creation of reciprocal relationships of trust. Moreover, there is evidence suggesting that, in certain cases, mechanism-design solutions may backfire (Bowles 2008). At the very least game theory provides a precise analytical framework that can be used to recognize the sub-optimality of certain conservation situations relative to a well-defined set of assumptions, while pointing towards possible solutions.

### 4.1.2. OUTLINE AND OVERVIEW

This chapter first presents three case studies of conservation dilemmas modeled using game theory in section 2. Section 3 discusses the empirical basis for the feasibility of informal solutions to these types of dilemmas that rely on members of the community to deliberate and enforce norms. Results from behavioral economics are marshaled to support the feasibility of these types of solutions, and to argue that in some situations, traditional solutions appealing to material incentives may backfire. Section 4 concludes by responding to Norton's (2005) critique of the use of game theory in environmental decision contexts.

---

worse off. An outcome is *Pareto-inefficient* if there exists some other outcome such that at least one agent is made better off while no agent is made worse off.

[68] Mechanism design, often called "reverse game theory," is the theory of designing games (rules, contracts, structures of incentives, etc.) such that a particular result is achieved (Myerson 2008).

## 4.2. Conservation Dilemmas

### 4.2.1. Wild Dogs and Local Villagers in South Africa

#### 4.2.1.1. Background

In South Africa, endangered carnivorous wild dogs (*Lycaon pictus*) were re-introduced into conservation areas in 1980–1981, and again in 1997 and in the early 2000s (Creel and Creel 2002, Gusset et al. 2008). The conservation plan analyzed here involved re-introduction of the species to the 900 km$^2$ Hluhluwe-iMfolozi Park in eastern South Africa (Maddock 1999), notable for attracting many South African and international visitors, primarily ecotourists. The park contained numerous large carnivores, including spotted hyaena (*Crocuta crocuta*), black-backed jackal (*Canis mesomelas*), cheetah (*Acinonyx jubatus*), lion (*Panthera leo*), and leopard (*Panthera pardus*). Conservation proponents intended to create meta–populations of *Lycaon pictus* that would be managed with occasional translocation between sub–populations to facilitate gene flow (Lindsay et al. 2004). By 2004, after more than 20 years of sporadic conservation measures, it was reported that the park itself supported nearly 50 dogs living in six packs, with an unknown number living in the surrounding unprotected areas.

Both conservation proponents, concerned that only about 6,000 individuals of this species remained in the wild, and the ecotourism industry, which found that tourists rated seeing the wild dogs quite highly, had an interest in promoting the re- introduction and translocation policies (Lindsay et al. 2005). However, rural herders and game farmers had an interest in the safety of their livestock or game populations, and many of them adopted a policy of killing wild dogs and other carnivores that escaped from conservation areas.

Although the local farmers, herders, and gamekeepers on private land, as well as Zulu villagers on communal land, were partly protected by the Hluhluwe-iMfolozi Park's electric perimeter fence, many of the large carnivores, especially the wild dogs, were known to escape from the park. Local community members held wild dogs responsible for roughly 15% of the annual livestock loss (Gusset et al. 2008). In response, conservation proponents accompanied the re-introduction and translocation policies with a public–relations campaign and a conservation education program for surrounding communities from 1999 to 2000. Results were assessed for program effectiveness in 2003. While ecotourists consistently reported positive attitudes toward seeing the wild dogs, and were willing to pay up to $150 for a chance to see them, villagers' attitudes toward the conservation program became more negative between 1999 and 2003. Furthermore, among those with limited educational background, misconceptions about the wild dogs and the goals of biological conservation were found to be widespread, and escaped dogs continued to be occasionally killed despite legal protection.

#### 4.2.1.2. Game Theoretical Analysis

The game represented in Table 1, which has the structure of a Prisoner's Dilemma (PD), can be used to represent the conflict between the conservation proponents (row) and local herders (column). Conservation proponents and the ecotourism industry are treated as one agent, A, because of their common shared interest; in the analysis below they will be referred to as conservation proponents. Each action available to A corresponds to a row of Table 1: these are to continue the re-location and translocation policy (T) or not do so (-T). Similarly, the herders and game farmers are treated as one agent, B, and will be referred to jointly as local herders. The actions available to B correspond to the columns: these are to have a policy of killing escaped dogs (K) or not do so (-K). The numbers represent ordinal rankings of the outcomes, where 1 is the best outcome, 2 is the next best outcome, and so on, and are given <Row, Column> with the first entry indicating the rank for A and the second the rank for B. The standard assumptions of one-stage games are applicable: each agent has full knowledge of its preference structure and is a competent maximizer over its own preference ordering.

Table 4.1. Two-Agent Game with Pareto-inefficient Nash Equilibrium

| | K | -K |
|---|---|---|
| T | 3,3 | 1,4 |
| -T | 4,1 | 2,2 |

Obviously, the best outcome for B is <-T, K>, while the best outcome for *A* is <T,-K>. The worst outcome for B is clearly <T, -K>, assuming the wild dogs are responsible for significant livestock loss. The worst outcome for A is <-T, K>, since no conservation translocation is pursued while B's policy threatens the feasibility of future conservation programs. The second– and third–best outcomes for A are <-T , -K> and <T, K>, respectively, on the assumption that the translocation policy comes at significant cost, and if killing takes place the cost of the translocation program would not be worth the little conservation value it would generate. The second- and third-best outcomes for B are <-T, -K> and <T, K>, respectively, on the assumption that without a translocation policy fewer wild carnivores threaten their livestock, while killing the escaped wild dogs is itself costly.

For A, T is preferred to -T, since whatever B's policy, the outcomes in which translocation policies are pursued are ranked higher: 3 as opposed to 4 and 1 as opposed to 2. The same reasoning on preferences shows, for B, K is preferred to -K. The unique Nash equilibrium is thus <T, K>, since neither agent can do better by unilaterally deviating from the strategy already being followed. (In pure strategies each of the agents only uses one of the available options and does not mix them in some proportion.) This equilibrium outcome, however, is Pareto-inefficient, since <-T, -K> is ranked 2 for both agents as opposed to 3. While <-T, -K> is not the unique Pareto-efficient solution, since <-T, K> and <T, -K> are most preferred by *B* and *A*, respectively, and ipso facto Pareto-efficient, these latter two outcomes are unattractive solutions as they are the least preferred by some agent.

#### 4.2.1.3. Discussion

Gusset et. al. (2008) have analyzed this conflict in some detail but did not note its relation to the Prisoner's Dilemma. Besides documenting the existence of the conflict between conservation proponents and local herders, they provided insight into possible solutions that prioritize conservation (and thus assume that T is necessarily preferred to -T). These solutions include continuing programs of conservation education, compensation measures for livestock loss, and participatory management policies (Ogada et al. 2003). Modeling the situation as a game provides additional insight. Any conservation-prioritizing solution to the conflict must either alter the payoffs for the local stakeholders (the herders), by deincentivizing K or incentivizing -K, via conventional mechanism-design solutions involving (effective) law enforcement and/or financial incentives, or else directly alter the preferences of the locals, which was presumably the goal of conservation education. Gusset et. al. noted that most of the locals had generally negative views of wild dogs. This suggests that improved husbandry practices combined with conservation education may be the most cost-effective solution.

### 4.2.2. Raptors and Red Grouse in Scotland

#### 4.2.2.1. Background

In Britain, in the 1990s, the relationship between raptors and their avian prey emerged as one of the more contentious issues in discussions of natural habitat conservation and management (Thirgood et al. 2000, Thompson et al. 1995). Whereas many raptor species' populations had begun to recover from their earlier pesticide-induced low levels of the 1970s, their prey species' populations were often in decline. Thirgood et al. (2000) reviewed how this conflict was being played out in the case of the Hen Harrier (*Circus cyaneus*) and Red Grouse (*Lagopus lagopus scotius*) on heather moorlands dominated by Ling Heather (*Calluna vulgaris*). The present distribution of these heather moorlands is largely limited to Britain and Ireland with smaller areas elsewhere in Europe. Consequently, in Britain, retention of these moorlands was considered to be of high conservation priority.

Heather moorlands supported unusually high populations of Red Grouse. Though many other bird species also utilized this habitat, Red Grouse was the only species entirely restricted to it. However, for most of those who wanted to preserve the moorlands, their retention was motivated not by concern for the ultimate survival of this species but, rather, because Red Grouse shooting was central to local economies. The primary aim of Red Grouse management had always been to maximize the number of individuals available for shooting every fall. Gamekeepers attempted to achieve this aim through the control of parasites and predators of Red Grouse populations. Among birds, three raptor species were among the implicated predators: the Hen Harrier, the Golden Eagle (*Aquila chrysaetos*), and the Peregrine Falcon (*Falco peregrinus*). The most important of these (by far) was the Hen Harrier. Hen Harriers, in turn, were prey for Golden Eagles. Though Golden Eagles presumably also preyed on Red Grouse, their role in controlling grouse populations was presumed to be minor compared to that of Hen Harriers.

Thirgood et al.'s review of the raptor-grouse conflict identified three potential and actual actions that would affect conservation prospects of the three species:

K: Hen Harriers could be culled to control their populations. The expected result would be increases in Red Grouse populations and the economic benefits associated with it. Culling was already taking place through hunting, which, though technically illegal, was nevertheless apparently widely practiced.

D: Diversionary feeding (e.g. carrion) could be introduced for Hen Harriers. This was believed to be able to decrease the predation pressure on Red Grouse though not to the same extent as K. It is assumed in this analysis that this action would to some extent benefit Hen Harrier populations at least so long as culling (K) is not undertaken. If culling were introduced, it is likely that D would have very little, if any, effect.

I: Golden Eagles could be introduced into Hen Harrier habitat. It is assumed (as was very likely) that the benefit to Red Grouse due to Golden Eagle predation of Hen Harriers outweighs the loss due to predation of the Red Grouse. (The analysis below will make the same assumption.)

It is next shown that each of these potential actions falls under the jurisdiction of a unique agent (an interest group consisting of an easily distinguished set of stakeholders).

#### 4.2.2.2. Agents and Goals

From Thirgood et al.'s description, there were many stakeholders involved in the dispute. However, it turns out that these varied stakeholders can be naturally organized into interest groups, each coupled to one of the actions identified above. The principle used for this grouping is that members of each group strongly share interest in some action that the group would encourage and different groups disagree on what that action is.

Thus, each of the following three interest groups will be treated as a single agent in the game theoretic analysis below:

A1: Gamekeepers and others who were economically dependent on Red Grouse hunting and wanted their populations to be as large as possible so as to maximize profits from hunting. It is unproblematic to expect that A1 would have control over K since it is in its interest to cull Hen Harriers.

A2: Hen Harrier conservationists who were concerned primarily with the welfare of that species, in part because they had once disappeared from all of Britain with the exception of the Scottish islands of Orkney and the Hebrides. A2 would presumably have almost complete control over D, since that action has some potential to help the Hen Harrier population at least when culling does not occur.

A3: Golden Eagle conservationists who were similarly primarily concerned with the welfare of that species. Presumably A3 would have sole control over I because of its expense, and in spite of probable reservations of A2, because in carrying out I, A3 would have at least some support from A1.

In one respect this characterization of the interest groups may be slightly artificial since Thirgood et al. do not distinguish Hen Harrier and Golden Eagle conservationists quite as sharply. However, it is useful to distinguish them because of the potential for conflict between Hen Harrier and Golden Eagle conservation due to the former being a potential prey of the latter, a problem that Thirgood et al. do note.

#### 4.2.2.3. Preference Analysis

Table 2 below shows the rank order of the preferences of the agents for each of the eight possible sets of three actions that can be taken by the agents. These form the set of alternatives in this decision analysis with each action, K, D, and I (performing it or not) being an available option for the agent associated with that action. This means that A1 can only choose between K and -K, A2 between D and -D, and A3 between I and -I. An outcome consists of one action each from each of the three agents, and the complete preference structure consists of an ordinal ranking of the entire outcome set by each of the agents, ties allowed, with 1 being the most preferred, and so on.

Table 4.2. Preference Structure

| Outcome | A1 | A2 | A3 |
|---|---|---|---|
| K, D, I | 1 | 7 | 3 |
| K, D, -I | 2 | 5 | 5 |
| K, -D, I | 3 | 6 | 3 |
| K, -D, -I | 4 | 4 | 5 |
| -K, D, I | 3 | 2 | 1 |
| -K, D, -I | 5 | 1 | 4 |
| -K, -D, I | 5 | 3 | 2 |
| -K, -D, -I | 6 | 2 | 4 |

For A1, clearly (K, D, I) is the best outcome (that is, it has rank 1), because each of these actions benefits Red Grouse. Assuming that Red Grouse predation by Golden

Eagles does happen to some extent (though it is not as serious as culling), the next best outcome is (K, D, -I). Both (K, -D, I) and (-K, D, I) are ranked 3, assuming that the combined effect of diversionary feeding and predation and the crucial fact that no effort is expended by A1 in the latter case cancels out the effect of culling in the former case. Since culling Hen Harriers is potentially a very effective way to reduce Red Grouse mortality (K, -D, -I) is ranked as 4. There is probably not much to distinguish (-K, D, -I) and (-K, -D, I)—these are both ranked as 5. (-K, -D, -I) is the worst because no action at all is taken to augment Red Grouse populations.

Agent A2's concerns are limited to Hen Harriers. Diversionary feeding, along with no culling and no predation, that is, (-K, D, -I), is the best option. Keeping the other two acts as they are, while not introducing diversionary feeding, that is, (-K, -D, -I) comes in at 2 as, from the same type of reasoning, does (-K, D, I). By losing diversionary feeding, (-K, -D, I) gets rank 3. It is assumed that when culling (K) occurs, diversionary feeding (D) does little to augment Hen Harrier populations, but predation (I) still has a small negative effect on them. Moreover, A2 presumably does not want to waste effort in performing D if it does not help Hen Harriers. Thus, taking wasted effort into account, (K, -D, -I) is given rank 4, (K, D, -I) rank 5, and (K, -D, I) rank 6. The situation is worst when both culling and predation occur, and A2 also wastes effort, that is, (K, D, I).

Turning to A3, the best outcome for Golden Eagles is clearly (-K, D, I), when the species is being introduced in Hen Harrier habitat and the main prey species is being encouraged to grow by no culling and diversionary feeding. For Golden Eagles, the outcome is only slightly worse if Hen Harriers lose diversionary feeding: (-K, -D, I) has rank 2. Beyond these two cases, assuming that diversionary feeding is not very important for Hen Harrier populations, the ranks A3 gives will be neutral with respect to D and -D.

Both (K, D, I) and (K, -D, I) will be ranked 3. Next come (-K, D, -I) and (-K, -D, -I). The worst scenarios are (K, D, -I) and (K, -D, -I).

#### 4.2.2.4. Game Theoretical Analysis

The decision scenario discussed above can be modeled as a 3-agent game with each agent, A1, A2, and A3 having control over one action: K, D, and I, respectively. As noted earlier, this is a simplifying but plausible assumption in this context. Agents' preferences over the eight possible outcomes are enumerated in Table 2.

This game will be analyzed to determine which outcomes, if any, are Nash equilibria and which are Pareto-efficient. In Table 3, it is shown that there is a unique Nash equilibrium, which is the outcome, (K, -D, I). In Table 4 it is then shown that there are four Pareto-efficient outcomes, (K, D, I), (K, D, -I), (-K, D, I), and (-K, D, -I). In other words, the Nash equilibrium, (K, -D, I), is a Pareto-inefficient outcome. In fact, it is Pareto-inefficient relative to (-K, D, I), which would leave no agent worse off and A2 and A3 better off.

Table 4.3. Nash Equilibrium Analysis

| Outcome | Stability | Analysis |
|---|---|---|
| K, D, I | Unstable | A2 can unilaterally deviate to (K, -D, I), ranked 6 instead of 7. |
| K, D, -I | Unstable | A2 can unilaterally deviate to (K, -D, -I), ranked 4 instead of 5. A3 can unilaterally deviate to (K, D, I), ranked 3 instead of 5. |
| K, -D, I | Stable | No agent has an incentive to deviate unilaterally. This is the Nash Equilibrium. Consider each agent's possible unilateral deviations. A deviation by A1 would result in (-K, -D, I), ranked worse, 5, instead of 3. A2's deviating would result in (K, D, I), ranked worse, 7, instead of 6. A deviation by A3 would result in (K, -D, -I), ranked worse, 5, instead of 3. |
| K, -D, -I | Unstable | A3 can unilaterally deviate to (K, -D, I), ranked 3 instead of 5. |
| -K, D, I | Unstable | A1 can unilaterally deviate to (K, D, I), ranked 1 instead of 3. |
| -K, D, -I | Unstable | A1 can unilaterally deviate to (K, D, -I), ranked 2 instead of 5. A2 can also unilaterally deviate to (K, -D, I), ranked 1 instead of 4. |
| -K, -D, I | Unstable | A1 can unilaterally deviate to (K, -D, I), ranked 3 instead of 5. A2 can also unilaterally deviate to (-K, D, I), ranked 2 instead of 3. |
| -K, -D, -I | Unstable | A1 can unilaterally deviate to (K, -D, -I), ranked 4 instead of 6. A2 can also unilaterally deviate to (-K, D, -I), ranked 1 instead of 2. A3 can also unilaterally deviate to (-K, D, I), ranked 2 instead of 4. |

Table 4.4. Pareto-efficiency Analysis (I: Pareto-inefficient; E: Pareto-efficient)

| Outcome | Efficiency | Analysis |
|---|---|---|
| K, D, I | E | Since this is A1's unique best outcome, switching outcomes could only make other agents better off by making A1 worse off. |
| K, D, -I | E | A1 and A3 would be better off by switching to (K, D, I), but this would make A2 worse off. Any other switch would make A1 worse off. |
| K, -D, I | I | A switch to (-K, D, I) would make A2 and A3 better off, while leaving A1 with the same rank of 3. Note that this is the unique Nash equilibrium. |
| K, -D, -I | I | All agents could be made better off by switching to (-K, D, I): A1 from rank 4 to 3, A2 from rank 4 to 2, and A3 from rank 5 to 1. |
| -K, D, I | E | It is A3's unique best outcome, so by the same reasoning as above, it is trivially Pareto-efficient. |
| -K, D, -I | E | This is A2's best outcome. |
| -K, -D, I | I | A switch to (-K, D, I) would make everyone better off: A1 from rank 5 to 3, A2 from rank 3 to 2, and A3 from rank 2 to 1. |
| -K, -D, -I | I | A switch to (-K, D, I) would make everyone at least as well off: A1 from rank 6 to 3, A2 stays at rank 2, and A3 from rank 4 to 1. |

#### 4.2.2.5. Discussion: Raptors and Red Grouse

The assumptions about group decisions that are made in computing the set of Pareto-efficient outcomes were minimal. It is only assumed that the outcomes have a complete ranking with ties allowed (a complete weak ordering) on the basis of each agent's preferences. There was no assumption made about whether the outcomes can be given quantitative (cardinal) values. If there were more information available on agents' preferences, more structure can be given to the set of Pareto-efficient outcomes.

An important limitation of this analysis is the restriction to pure strategies: agents do not have the option of mixed strategies in which they sometimes carry out one action and sometimes do not. Moreover, a problem remains: there are four Pareto-efficient outcomes: (K, D, I), (K, D, -I), (-K, D, I), and (-K, D, -I) and only one of these can be implemented. The set of Pareto-efficient outcomes may have to be analyzed further to come up with a credible policy recommendation. There are at least two options available at this stage.

Additional assumptions about agents' preferences can be introduced to compound them to produce unique results. Methods range from simple voting to aggregating individual utility functions into a group utility function. None of these methods is devoid of conceptual problems. For example, using the kind of method discussed in chapter 2 assumes that tradeoffs between the utilities of individuals are acceptable. This assumption may not hold, especially when people believe that control of their land, for example, is a non-negotiable right.

Sorting out the Pareto-efficient alternatives may be handed over to a deliberative process in which the agents discuss these outcomes. Though not immune to the charge of being *ad hoc*, some criteria that may be used have some reasonable intuitive support. For instance, any extremal outcome (an outcome that is the most preferred by any of the agents) will always be Pareto-efficient no matter how poorly it is ranked by all other agents. It may, therefore, be reasonable to drop most of such extremal outcomes: in the case study of this paper, (K, D, I), (-K, D, I), and (-K, D, -I) would be dropped leaving only (K, D, -I) as a policy recommendation. Another method may be to deliberate on the values of all agents. In the case study here, it is reasonable to suppose that A2 and A3 may have moral scruples about killing animals. Thus, they may want to drop (K, D, I) and (K, D, -I) and then agree to choose (-K, D, I) over (-K, D, -I) because that is in accord with A1's preferences.

### 4.2.3. The n-Agent Dilemma: Reef Fishermen in the Philippines

#### 4.2.3.1. Background

Coral reefs, especially those in the southern Philippines and central Indonesia, are widely regarded as biodiversity “hotspots” of high conservation priority (Roberts et al. 2002). These rich marine ecosystems are home to hundreds of thousands of fish, bivalve, gastropod, cephalopod, crustacean, echinoderm, algae, and other species, many of which are typically micro-endemics. While human activities on land contribute to reef degradation via the downstream effects of agricultural and logging activities, industrial run-offs and other pollutants, in the marine arena, overfishing and destructive fishing techniques (e.g., those using improvised explosives or sodium cyanide) have also been centrally implicated in reef destruction (Roberts et al. 2002). These reefs are often vital to local economies. In the Philippines, for example, over-crowded coral fisheries support an economic livelihood for over a million fishers (White et al. 2000). The destructive ecological effects of overfishing on coral reefs are well documented. Two examples will help set the context (McManus 1997): (i) in the Philippine coral reef system of Bolinao, overfishing led to near extinction for the sea urchin (*Tripneustis gratilla*), which had been formerly quite abundant in the reef’s seagrass beds; (ii) in Kenya reefs were threatened by overfishing because the removal of high-level predators led to a dramatic increase in populations of drupellid snails which feed on coral.

According to McManus et al. (1997), roughly 350 marine species from the 40 $km^2$ Bolinao reef area are sold in local markets. In spite of the practice being banned in 1979, fishers continue to use explosive fishing techniques and have a strong financial incentive to do so: dangerous homemade bombs are cheap to produce at U.S. $1-2 and can generate a catch worth U.S. $15-40 while the average fisher, using non-destructive techniques, generates only about U.S. $1 a day. They report that informal surveys of the reef area in the mid-1980s showed that 60% of scleractinian coral was dead, much apparently due to fishing with explosives. Furthermore, their simple models indicated that fishing with explosives may have reduced the growth capacity of scleractinian coral by a third or more, with predictably negative effects for biodiversity.

Game theoretical analyses have been used in many analyses of fishing policies (Sumaila 1999). The open-access version of the n-agent game described below corresponds to the classic "tragedy of the commons." While the analysis is simple, it captures the dynamics of overfishing in coral reefs in Bolinao where the resource is over-exploited because of no clear established rights of use. The Nash equilibrium outcome of collective over-exploitation of fish and the use of destructive fishing techniques is both economically undesirable (because of Pareto-inefficiency), as well as a major threat to healthy reefs and, thus, to sustainability and the conservation of biodiversity. However, it is then shown that, even in a closed-access n-agent gave, there can be a conflict between resource management policies based on "maximum sustainable yield" (MSY) and biological conservation. In this case, due to the ecological interactions between the exploited fish and other reef species of high conservation priority, MSY harvest levels for the exploited species may lead to a decline of other species targeted for conservation as important components of biodiversity. The analysis assumes that, in the long run, this trend leads to a decline in the exploited species because of mutualistic interactions—however, this part of the analysis should be regarded as a conceptual exercise rather than an exploration of the data.

### 4.2.3.2. Game Theoretical Analysis.

The Gordon–Schaefer model of open–access fisheries (Gordon 1953, 1954), as well as Hardin's (1968) less formal ''tragedy of the commons'' model of common pool resources, predict overfishing when individual or collective access/property rights to fisheries are ill-defined. The "bionomic equilibrium" is the point at which the population is so depleted that even minimal harvesting effort is not worth the expected return (Ludwig 2001).

This situation can also be represented as an n-agent PD, with the payoffs for agents along the rows as given in Table 5. The payoffs are symmetric for all agents, in the sense that all agents find themselves in the situation described by the payoff matrix. The non-cooperative action, D, is to harvest as much as possible now (or, in the case of overfishing in Philippine reefs, use destructive fishing methods like dynamite or cyanide). $N_D$ denotes the number of agents who play D, the non-cooperative harvesting effort, and $t \leq N$ is some threshold value (''tipping point'') such that, where t or more agents defect, the outcome shifts from the left to the right column: the common–pool resource is overexploited and fishing is not worth the effort. We assume the cooperative action C is to restrain harvesting effort to a level such that, if $N_D < t$, the population is sustainable over time.

Table 4.5. Open-access n-agent Game. Agents: n fishers in an open-access fishery. Strategies: D: Harvest as much as possible now; C: restrain harvesting effort to maximum sustainable yield levels. $N_D$: number of agents who play D. t is tipping point where harvesting effort exceeds maximum sustainable yield levels. It is assumed that $T > R > P > S$ for each fisher.

| | $N_D < t$ | $N_D \geq t$ |
|---|---|---|
| C | R | S |
| D | T | P |

For the preference structure, we assume T > R > P > S. This means that the worst case for each agents is to restrain harvesting effort while others overexploit the fish. The best case is for an insignificant number of others (< t) to defect while agents harvests as much as possible. The second-best case for agents is to cooperate while all (or a significant number) of others cooperate by restraining harvesting effort. The third-best case for an agent is to defect, achieving a short-term gain while the fish population reaches bionomic equilibrium: enough agents defect such that the population is over-exploited in a short time. An agent does better by defecting no matter what the others do, since T > R and P > S. The Nash equilibrium solution of this game is the situation in which all defect, and the fish are overexploited. This is each agent's third-best outcome, whereas if everyone cooperated they would have achieved their second-best outcomes, and the exploited population of fish would persist at a sustainable level.

In this case, biodiversity values and economic values both prescribe conservation action. Economically, the open-access Nash equilibrium is inferior to the cooperative outcome for every agent: the latter is strictly preferred to the former by every agent. Furthermore, the destructive fishing techniques and overfishing that characterize the open-access equilibrium clearly threaten reef integrity and biodiversity.

However, a second n-agent PD may arise that pits economic and conservation values against one another in the short term. Consider the situation in which the open-access problem (the "tragedy of the commons") for some reef fishery has been solved by privatization, government control, or community management, such that a resource management plan for "maximum sustainable yield" (MSY) of the fish has been instituted. We still assume there are a number of agents extracting fish, but in this game the Nash equilibrium is for the agents to restrain their harvesting effort to the MSY level.

Crucially, we make an assumption about the ecological interactions between the exploited fish and the surrounding reef ecosystem: the MSY harvesting effort will, over time, lead to a slow decline in some endemic species of high conservation priority. As the population of this second species declines, the population of the exploited species will as well, such that the "MSY" harvesting effort is actually unsustainable. (For a partial justification of this assumption, see the cases described by Redford and Feinsinger 2001.)

This second, parallel n-agent PD is represented in Table 6. In the open-access n-agent game, we assumed that some policy similar to MSY harvesting was the "cooperative" option. In the closed-access case, the cooperative action will be denoted by BCE, for biodiversity conservation effort, and the non-cooperative option is the more intensive MSY harvesting, denoted by MSY. Otherwise, the preference structure is exactly parallel. The Nash equilibrium solution sustains the fish species in the short term, but as the second species slowly declines, in the long run, catches of the economically valuable fish decline in turn. Thus the Pareto-efficient solution involves each agent restraining harvesting effort beyond the short-term MSY point.

Table 4.6. Closed-access n-agent Game. Agents: n fishers in a closed-access fishery. Strategies: MSY: harvest at maximum sustainable yield levels. NM: number of agents who play MSY; t is tipping point where harvesting effort leads to eventual decline in yield due to ecological interaction with species of conservation value. It is assumed that T > R > P > S for each fisher.

| | $N_M < t$ | $N_M \geq t$ |
|---|---|---|
| BCE | R | S |
| MSY | T | P |

#### 4.2.3.3. Discussion: Fish and Corals in Southeast Asia.

Overfishing in coral reefs is both an economic and biodiversity conservation issue, especially in cases like that described by White et al. (2000), in which fish levels in some areas of the Philippines have dropped below those necessary to sustain healthy coral reefs. They report that while healthy reefs can sustainably produce 20 ton/km$^2$/year of edible products, reefs degraded due to overfishing or cyanide use produce less than 4 ton/km$^2$/year. Other economic benefits attaching to the preservation of biodiversity in reefs include revenue from tourism: reef diving, tour fees, etc.

Admittedly, the closed-access n-agent PD is only a speculative ecological model, but it brings into focus the need for conservation and resource management planners to take long-term ecological interactions into account in assessing solutions to conservation-relevant economic conflicts. In particular, it shows that appeal to theories of sustainable exploitation may at best produce short-term Pareto-efficient outcomes.

The formal mechanism-design solution to the n-agent PD alters the behavior of each agent via material incentives or the threat of punishment by defining clear use rights. It is well-known that enforceable government ownership, group ownership, or individual ownership can go a long way towards preventing over-exploitation of resources (Ostrom et al. 1999). Even in closed-access fisheries, however, further regulations and incentives may be necessary to ensure sustainability when multiple users compete (OECD 1997). While such formal solutions can be effective, two comments are in order. First, as discussed in the next section, resource users can and do develop informal networks of trust and reciprocity norms that can solve open access dilemmas (Pinkerton 1989). Second, appealing solely to agents' self-regarding preferences can be counterproductive. In the next section evidence is presented that suggests limitations of such narrow mechanism-design solutions.

## 4.3. Cooperation and The Limits of Mechanism Design

### 4.3.1. Cooperation in Conservation Dilemmas

What should happen when Nash equilibria are Pareto-inefficient, like in the cases presented above? It is easy to say that agents should cooperate, and very broadly, there are two ways this could happen. The first is via the intervention of a third party, whether a government or a hired mediator, who alters the incentives that led to the dilemma in the first place. A government or hired third party (Ostrom 1990) could police the reef fishery, for example, restricting the access of fishers to prevent overfishing in the open-access n-agent dilemma above. This often-preferred type of solution will be called the 'mechanism design' solution, after the branch of game theory that studies the implementation of outcomes via the design of rules and the alteration of incentives (Myerson 2008). These types of solutions assume that the government or third party can be trusted to provide adequate incentives for cooperative behavior in a way that does not lead to inefficiencies (e.g. enforcement that is too costly to be worth the investment) or other problems (e.g. abuses of power).[69]

In the absence of policies that provide third party enforcement of cooperative strategies, willingness to cooperate in repeated interactions depends on the level of trust, more specifically, the degree of confidence an agent has that another agent will not unilaterally change strategy. There are various ways that such trust could be built. The obvious suggestion is more discussion and deliberation and, especially, repeated interactions where agents have an opportunity to build and enforce norms of cooperation themselves. In environmental decision contexts, given that few agents actively claim an

[69] For example, Ostrom et al. (1999) discuss the government-owned Chiregad irrigation system in Nepal, which replaced irrigation systems owned and managed by farmers themselves. Apparently the government's system, although technologically more sophisticated, did not take into account the local norms that had been used to allocate use rights in the older system, resulting in less use of the new system and lower agricultural productivity.

explicit desire to harm environmental goods and services, collective decision-making through deliberation is one plausible recommendation to begin articulating these norms (Fiorinio 1990).

In the South African case study, the pursuit of such deliberative strategies should presumably include conservation education and credible plans from conservation proponents to offset costs incurred by herders and game farmers due to predation by wild dogs. This recommendation is easier to make in this case because only two "agents" (groups) are involved. The situation in the British example is more complex, requiring reciprocal commitments between gamekeepers and the two classes of raptor conservationists. For instance, if each of the three agents agreed to drop a policy which is deemed best by only one agent, there would remain only one outcome, (K, D, -I), on which they would have to agree. In the case of overfishing on coral reefs in the Philippines, collective deliberation would presumably have to take place through public forums or community groups because of the number of agents that are involved. However, because the number of agents is so large, the opportunities for these agents to agree on and collectively enforce norms to avoid the dilemma in the absence of external enforcement is more difficult, especially since larger groups have trouble effectively monitoring each other and excluding outside agents who do not have use rights (Ostrom 1990).

Again, the contrast here is with the mechanism-design strategy for achieving Pareto-efficient outcomes that relies on alteration of material incentives by a third party. Usually, advocates of this approach assume agents are narrowly self-interested, and advocate policies that assume agents will act rationally according to self-interested preferences (Gintis 2000, Henrich et al. 2005). Bowles (2008) has noted this mechanism-design view makes strong and controversial foundational assumptions. Narrow self-interest is presumed to be the basis for social institutions and resulting institutional arrangements. Institutional policies are supposed to work best when designed for "knaves."

While policies based on third-party alteration of material incentives have some record of success, experimental results in psychology and behavioral economics show that there are significant limitations to narrow mechanism-design solutions that are particularly relevant in planning for environmental values, including biodiversity conservation. Firstly, a significant body of experimental work in behavioral game theory shows that humans have dispositions to initially cooperate and enforce norms of cooperation at a cost to themselves in dilemma-style games with Pareto-inefficient Nash equilibria. Secondly, independent streams of research in psychology and economics suggest that in certain types of situation mechanism-design solutions might backfire, by undermining the "moral sentiments" or other motivations that can contribute to cooperative behavior. In short, appeal to narrow (material) self-interest may "crowd out" other–regarding motives (Frey 1987). The next section discusses relevant results from behavioral game theory and the subsequent section discusses evidence for crowding out.

### 4.3.2. THE BEHAVIORAL GAME THEORY OF DILEMMAS

In this section I summarize evidence from behavioral game theory that suggests that people have social or other-regarding preferences, including preferences for reciprocity and the enforcement of norms. This evidence suggests that agents reciprocally cooperating over time can solve many dilemmas, especially where some agents are willing to "pay" to enforce norms that maintain such cooperation.

For each game, I give a description of the rules and monetary payoffs, the Nash equilibria, the observed results, and how they deviate from money-maximizing Nash equilibria. The experiments are designed to be anonymous to both the players and the experimenters: individuals do not know who they are playing against, and experimenters do not know which individuals play which strategies. The games also have finite horizons: they are either one-shot or finitely repeated. These are meant to control for reputation effects and indefinitely repeated game strategic play.

A crucial finding is that populations are heterogeneous: while individuals exhibiting cooperative behavior have been found to make up a significant proportion (sometimes a majority) of individuals in experiments, it is likewise found that a substantial proportion of individuals do indeed behave as rational money-maximizers (see, e.g. Fehr et al. 2002, Ostrom 2005).

#### 4.3.2.1. One-shot Prisoner's Dilemma (PD)

The game is shown in table 7 below, and shares the structure of the 2-agent conservation dilemma from South Africa presented above. Two players decide whether to cooperate (C) or defect (D), where the payoff to a lone cooperator (CD) is the lowest, followed by the payoff to mutual defection (DD), followed by the payoff to mutual cooperation (CC), followed by the payoff to lone defection (DC): CD < DD < CC < DC.[70]

Table 4.7. One-shot Prisoner's Dilemma. 'C' is the cooperative action, 'D' is defection. Payoffs are for row player; see text for preferences (same for row and column).

| | C | D |
|---|---|---|
| C | CC | CD |
| D | DC | DD |

[70] The game is played once, making strategically irrelevant the usual stipulation that 2CC > DC + CD.

Defection dominates cooperation, since whatever column player does, row player does better by defecting, and vice versa, since DC > CC and DD > CD. By this line of dominance reasoning, mutual defection is a Nash equilibrium, but the mutually cooperative outcome is Pareto-efficient, since CC > DD. In the laboratory, subjects cooperate about half the time. Since defection dominates cooperation, rational money-maximizers should always defect in the one-shot game. When given time to play the one-shot PD repeatedly, subjects generally converge to the Nash outcome (Sally 1995, Ledyard 1995).

#### 4.3.2.2. Public Goods Game

$N$ players decide how much to invest $c_i$ from their initial endowment $e_i$ in a "public good" fund, whose proceeds are shared equally by all, that pays $m$ per unit of investment, where $m < 1$ and $i$ is an index for the players. Payoffs for each individual i are thus: $P_i = e_i - c_i + m(\Sigma_k c_k)/N$. Since $m < 1$, there is no individual incentive to invest, however $mN > 1$, so everyone would benefit from full investment. The game is equivalent to an n-player prisoners' dilemma, and is played repeatedly over a series of $r$ rounds. The Nash equilibrium outcome in the one-shot game is zero investment, and the sub-game perfect equilibrium (where agents act in equilibrium for each sub-game) in the finitely repeated game predicts zero investment, although there is some controversy over the relevance of sub-game perfection in experimental settings (Camerer 2003).

In the laboratory, in the first few rounds of the finitely repeated game in industrialized populations, mean investment is about half of initial endowment, with a roughly bimodal distribution, i.e. roughly half of the individuals invest their entire endowment and half invest nothing (Camerer 2003, Fehr and Gächter 2000, Ledyard 1995). This leads to a sharp decline in investment over the course of the game.

#### 4.3.2.3. Public Goods Game with Targeted, Costly Punishment

This game is nearly identical to the public goods game as described above, but adds the following complexity. After each round, players may choose to pay to target particular players for punishment (e.g. those that defected). Parameters vary between experiments, but usually the cost of punishment is less than the amount taken away from the player punished. Again, the subgame-perfect Nash equilibrium outcome for the game iterated a finite number of times is zero investment. Similarly, the introduction of costly punishment creates a "second-order" public goods dilemma such that the sub-game-perfect Nash equilibrium outcome is that no one should pay to punish. The idea here is that the provision of punishment to sustain the Pareto-efficient outcome is itself a public good. In the laboratory, adding the opportunity for targeted, costly punishment can sustain investment at high, efficiency-promoting levels. Furthermore, subjects are willing to engage in third-party punishment of individuals that do not directly affect their income, and are also willing to punish non-contributors on the last round of play, when such punishment could not affect future contributions (Fehr and Gächter 2000, Fehr and Fischbacher 2004, Ledyard 1995).

#### 4.3.2.4. Public Goods Game with Targeted, Costly Punishment and Intergroup Conflict

Sääksvuori, Mappes, and Puurtinen (2011) recently devised a version of the public goods experiment looking at the effects of inter-group conflict. The rules are the same as the public-goods game described above, except the experimenters devised several new experimental treatments relating to competition between groups. The experimenters ran two symmetric competition treatments, one in which both groups were given the opportunity to punish non-contributors at individual cost, and one in which both groups were not given this opportunity. They also ran an asymmetric competition treatment, where one group was able to punish and another was not. In these group competition treatments, round-by-round payoffs for each group were public knowledge. The sub-game perfect Nash equilibrium outcome would result in no one contributing and no one punishing.

In the laboratory, levels of contribution and punishment in control treatments were similar to those described above. Punishment opportunity led to significantly higher levels of contribution to the public good in the asymmetric and symmetric group-competition treatments when payoffs were public knowledge. When groups were isolated, the effect was only marginally significant. Net payoffs to punishing groups were significantly higher in asymmetric group conflict. However, in the absence of group conflict there was no statistically significant difference between net payoffs for punishing and non-punishing groups (the costs of punishment canceling out the benefits of cooperation). If individuals were rational money-maximizers playing subgame-perfect Nash equilibria, they would neither contribute to the public good nor pay to punish, and the effects of group conflict are irrelevant to monetary payoff, so would not affect play.

### 4.3.3. The Limits of Mechanism Design

The folk theorem of repeated games (actually a class of theorems) states that rational agents can achieve Nash equilibrium outcomes that are not available as equilibria in one-shot games (e.g. mutual cooperation in the PD) when those games are (infinitely or indefinitely) repeated, since the opportunity for endogenous enforcement of such outcomes is available (Gintis 2009). For example, if the PD were repeated, the respondent may "punish" a non-cooperator by not cooperating in the next round. This behavior may furthermore be rational, since sufficiently forward-looking agents care about their income stream from future interactions: foregoing a positive but low payoff today may be rational if it causes the receipt of a larger payoff every day next week.

However, the results from behavioral game theory above show that people are often willing to cooperate reciprocally and enforce norms at a cost when faced with even single-shot or finitely iterated dilemmas. Thus mechanism-design solutions beyond processes of group deliberation may in some cases simply be unnecessary. Additionally, mechanism designers interested in designing institutional arrangements that result in Pareto-efficient Nash equilibria should take into account psychological complexities, lest their policy prescriptions backfire. Bowles (2008) discusses several reasons why such policies might backfire.

*Framing and informational effects:* Where cooperation is “framed” (in the psychologist’s sense) as required by regulation or law, and enforced, for example, by a fine, this may actually undermine cooperative behavior over time. It may be better to frame cooperation in the context of group decision-making amidst informal networks of communication, appealing to agents’ other-regarding motives (Cardenas et al. 2000). Further, material incentives may send a negative signal to the agents that can motivate defection, for instance, because they may indicate a lack of trust (Fehr and Rockenbach 2003).

*Learning effects:* Incentives may provide an environment in which agents “learn” to be more self-interested, and their preferences shift over time to become less other-regarding. For example, in the experiment of Falkinger et al. (2000), subjects who had played a public goods game in the presence of material incentives later played the game without incentives and contributed 26% less to the public good than subjects who had not played the game with incentives.

*Overdetermination*: A significant body of psychological research on "intrinsic" motivations suggests that, when agents are offered financial incentives for actions for which they are already intrinsically motivated (e.g., because they are pleasurable), intrinsic motivation may decrease significantly (Wiersma 1992, Deci et al. 1999, Gneezy and Rustichini 2000). Bowles and Hwang (2008) consider theoretical models where moral sentiments and material interests are either complements or substitutes, showing that a "naïve" social planner interested in maximizing aggregate utility who presupposes agents are purely self-interested may mistakenly advocate policies where incentives are too small or too large.

When these kinds of situations obtain, deliberative solutions that engage the moral sentiments and other-regarding motives of the agents are likely to better achieve the goals envisioned by them. Encouraging agents to communicate and reach agreements on behavior, for instance, through credible promises of future behavior to each other, may be sufficient. If agents are sure that others will not unilaterally change their actions, the Nash equilibrium of the static game becomes irrelevant.

## 4.4. Concluding Discussion: Norton's Critique

This chapter has been concerned with using game theory to identify dilemmas where each individual or group acting in their own interest results in an outcome that is worse for everyone. Thus game theory can serve a normative role for group decision-making. I conclude by considering Norton's (2005) critique of game theory in conservation contexts, arguing that his notion of an irreducibly communal good is not necessary to account for environmental values like those involved in the biological conservation dilemmas considered here. Furthermore his pragmatism about the formal tools of decision science suggests a weaker characterization of methodological individualism than he provides, which is not necessarily incompatible with methodological holism.

Norton's (2005, 232-242) main claim about the uses of game theory in conservation contexts is that contemporary economists' commitment to methodological individualism precludes an engagement with environmental values he calls "communal." Whether or not Norton's notion of an irreducibly communal good is necessary to characterize environmental decision contexts (and I give some reasons to doubt this below), I agree with Norton that the utility, scope, and limits of decision tools like game theory for environmental decision contexts are ultimately to be determined empirically.

According to Norton, environmental problems stem largely from conflicts between individual goods and communal goods. Individual goods are the fungible, relatively short-term goods that concern economists. A communal good is supposed to emerge at the spatiotemporal scale of a whole community over time, the multigenerational time scale wherein "a human community finds its proper niche in an ecological system" (2005, 241). Crucially, communal goods are not supposed to be reducible to an aggregation of goods accruing to individuals. Environmental values like common-pool resources (e.g. clean air and water) and biodiversity are exemplars. Norton claims, for example, that a *community* may be worse off due to the destruction of a pasture whether or not any individual in the community prefers to work the pasture as a herder. Without the pasture, members of the community lose a feasible *option* for productive work and an alternative way of life.

The example is suggestive but the notion of an irreducibly communal good is problematic, since someone interested in accommodating what Norton calls communal goods in terms of aggregations of individual goods have several plausible options. In Norton's example of the pasture, individuals could have (revealed, elicited, or constructed) preferences over which lifestyle options should remain feasible for future generations. If no individual in the current generation or in future generations prefers to be a herder, perhaps due to cultural transmission of non-herder values, it is far from clear whether this community is worse off if the pasture is destroyed. Preferences over which options remain feasible whether or not future individuals pursue those options could be elicited, as individuals in this community could aggregate their preferences by voting to indefinitely conserve the pasture,[71] or revealed via market behavior or via community action to conserve the pasture. If individuals were not given an opportunity to express such preferences and the pasture were destroyed, the community may be said to be worse off because a large enough majority or perhaps plurality of individuals would have preferred to indefinitely conserve the pasture. Here we do not appeal to communal goods, but an aggregation of individual (elicited or revealed) preferences to support the claim that the hypothesized community is worse off.

Two other options are open which avoid the notion of an irreducibly communal good. Norton could appeal to his own notion of a considered preference (Norton 1984), arguing that while the community could be better off according to an aggregation of their "felt" or revealed preferences if the pasture were destroyed (since no one currently wants to be a herder while they believe they would immediately benefit from conversion of the pasture to an alternative use), they are also worse off according to their considered preferences, on the assumption that in ideal circumstances most individuals would reflectively endorse the preference for the herder lifestyle option being left open. Perhaps less plausibly, Norton could appeal to a purely objective notion of well-being that leaves the economist's subjectivist preference-based conception of value behind, claiming that

[71] The problem of aggregating individual preferences in voting systems is non-trivial, however the impossibility results of Arrow's theorem do not apply to binary choices, e.g. a choice between conserving and not conserving a pasture (Arrow 1951).

individuals are better off the more options they have, whether or not they ever wish to pursue them.

Philosophically, it is contemporary economists' supposed commitment to methodological individualism that Norton finds problematic, since it seems to preclude communal goods by definitional fiat. Methodological individualism is a research tradition with a long and complex history in philosophy and social science (Udehn 2001), arguably beginning with Hobbes's account of the social contract. It is unclear whether methodological individualists' opposition to so-called "organicism" or "holism" can be cashed out non-trivially. Norton glosses methodological individualism as the claim, "the legitimate object of social scientific study is ultimately the individual human being," (2005, 238) but his interest is particularly in economists' definition of social or communal goods as ultimately composed of (reducible to) aggregations of goods that accrue to individuals. His argument is that economists are involved in a kind of semantic sleight of hand by defining out of existence any kind of good that cannot in principle be reduced to individual goods.

An economist might argue, perhaps along the lines given above, that goods that emerge on a multigenerational scale can in principle be reduced to goods accruing to individuals over time. I do not think that it is a non-starter: many of the agents in the game theoretical representations of the conservation dilemmas offered above can plausibly be said to have preferences over multigenerational outcomes. For example, the agents in the case from Scotland have preferences for the persistence of bird populations beyond the current generation; in the case from the Philippines, presumably fishers have some preference for the persistence of the reef that gives them a livelihood. Of course, the agents' revealed, elicited, and constructed (or considered) preferences might come apart, as agents may act myopically in their short-term interest at the expense of future generations, but would not endorse these revealed preferences upon reflection. Thus any particular model of these agents' multigenerational preferences may be descriptively adequate but normatively inadequate, or vice versa.

Whether or not Norton's notion of an irreducibly communal good is coherent, it is not clear that game theory need be methodologically individualist in Norton's sense at all: the two- and three- agent examples above illustrate that *groups* can be modeled as single agents when interests are shared. Groups of individuals (firms, committees, government agencies, etc.) that interact in ways that allow them to aggregate their preferences, for example by voting, may also be treated as agents with a group preference. A better view of methodological individualism, more compatible with Norton's pragmatism, casts it as a research strategy that is not necessarily incompatible with methodological holism. An *explanation* is methodologically individualist if it explains social phenomena by appealing to properties of individuals and their interactions, for example explaining the outcome of a social interaction as the equilibrium of a game between individual agents. In this sense, insofar as game theory can treat groups as agents in some circumstances, it is also a methodologically holist tool.

The examples and models considered here, and throughout the dissertation, support Norton's pragmatism about the formal tools of decision science. He writes that we should "consider the multiple techniques of decision analysts to be useful tools within a larger process that is based in experience and a commitment to experimentalism." (261) That is, whether and how game theory (or multi-attribute value theory) will prove useful for environmental decision contexts is an empirical question that cannot be settled a priori. The case studies presented above show that game theory can inform group decisions in biological conservation and other contexts in which environmental values are at stake. However, they do not support the claim that a notion of communal goods that are irreducible to individual goods is necessary to accommodate the goals of conservationists.

# Chapter 5: Ethical Dilemmas in Biological Conservation and the Limits of Decision Science

## 5.1. Introduction: Connecting Commensurability and Cooperation

While the first chapter discussed how values play a necessary but often obscure role in applied conservation science due to inductive and definitional risks, the previous three chapters showed how modeling tools from decision science, particularly multi-attribute value theory (MAVT) and game theory, could be used to make the valuations of decision-makers explicit, in order to construct common scales of value to analyze complex tradeoffs (chapters 2 and 3) and facilitate cooperation in "dilemma" situations (chapter 4). Since the focus here is on prescriptive or normative applications of the decision sciences, it is worth stepping back to discuss the limitations of such applications, particularly in the resolution of ethical dilemmas in biological conservation.

This chapter discusses ethical aspects of the problems of tradeoffs between multiple values and cooperation between multiple stakeholders, enumerating further assumptions that must hold for these kinds of decision theoretic arguments to be successful. Case studies and ethical reflection reveal that decision science cannot provide guidance without additional normative assumptions that: (1) identify legitimate stakeholders and rights holders, defining the circle of concern and allocating decision-making authority amongst groups with different interests; and (2) delimit reasonable tradeoffs, deciding which biological or ecological units should be protected and at what cost, and whose rights are non-negotiable constraints.

The especially difficult problem of non-negotiable constraints connects the problems of commensurability and cooperation: people who believe their rights or interests to be incommensurable with the interests of others, in the sense that they should lexicographically trump them or place hard constraints on decision-making, are generally more likely to act uncooperatively when those rights or interests are at stake. This applies to both conservation proponents, for example Wood (2000), who argues for the lexicographic priority of biodiversity protection in land use decisions,[72] and those who oppose conservation measures. Insofar as the herders in the South African wild dogs case from chapter 4 believe that their right to violently protect their herds from wild dog attack is simply non-negotiable, and thus will not accept some form of compensation in exchange for giving up this right, they will be less amenable to solutions that protect the endangered dogs.

While commitment to non-negotiable rights may lead to land use conflict in biological conservation scenarios, resulting uncooperative behavior may be ethically justified when vulnerable populations face mistreatment and injustice at the hands of conservationists and governments (Dowie 2009, Chapin 2004), or other resource users like extractive corporations (Okonta and Douglas 2003). On the other hand, accepting that at least some tradeoffs are unavoidable in land use decisions for conservation requires reasonable constraints on claims of incommensurable, non-negotiable rights.

This chapter proceeds as follows. Decision analysis for biological conservation often begins by identifying relevant decision-makers and stakeholders, and section 2 enumerates normative problems for stakeholder methodologies: determining ethically legitimate stakeholders, dealing with asymmetries of power and concerns for social justice, and weighing the rights and interests of various stakeholder groups. The last issue provides a natural transition to a discussion of the ethics of tradeoffs in section 3, which describes a case study of introduced North American beavers and local biodiversity in Navarino Island, Chile. The case study illustrates the weaknesses of the decision theoretic

[72] Studies such as Spash and Hanley (1995) and Rosenberger et al. (2004) show that it is possible to elicit lexicographic preferences for environmental goods from significant proportions of student populations.

methods described in earlier chapters for making normative arguments in ethically disputed territory. Section 4 concludes by discussing the value of decision science to logically clarify these disputes, but not resolve them without further normative inquiry.

## 5.2. Normative Problems for Stakeholder Methodologies

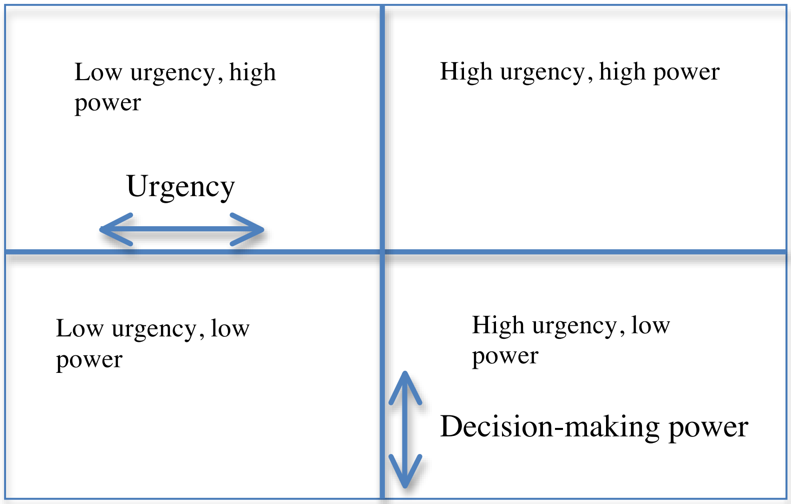
### 5.2.1. Legitimate Stakeholders and Social Justice

Decision analysis of the sort discussed in chapters 2-4 begins by explicitly or implicitly identifying relevant decision-makers and agents affected by outcomes of decisions. Freeman's (1984) influential, very broad definition of 'stakeholder' is anyone who can affect or is affected by a decision. Two central normative issues for the stakeholder approach, which remain largely unresolved, are (1) what it takes for an individual or group to count as an ethically legitimate stakeholder; [73] and (2) how such methodologies should account for asymmetries in power between stakeholder groups.

Many definitions of 'stakeholder' relevant to firms have been offered in the business and management literature.[74] They range from very inclusive (anyone affected by the firm's activities) to very restrictive (anyone with a legal obligation to or claim on the firm). Most analyses categorize stakeholders on at least two dimensions: decision-making power and potential costs and benefits borne or received ("urgency").[75] Figure 1 depicts this commonly used conceptual model, with urgency on the x-axis and decision-making power on the y-axis.

Figure 5.1. Stakeholder Categories in Terms of Urgency and Power

[73] This question is stressed by Sarkar (2012a).
[74] See, for example, Freeman (1984) and Donaldson and Preston (1995).
[75] For example, Mitchell et al. (1997).




This conceptual model is inadequate to answer the normative question of who counts as a legitimate stakeholder, however, since decision-making power itself may be (normatively) legitimate or illegitimate, and an interest group may use its power to pursue its own interests at the expense of other stakeholders. Furthermore, how the "urgency" of various claims is to be compared is a normative question: a real estate developer may have millions of dollars at stake in a land use decision, whereas conservationist groups may oppose the development on the grounds that it will threaten an endangered or endemic species.[76] Both groups would describe their interests as urgent, but how these stakeholders' claims should be weighed is a significant further question. These issues lead authors like Mitchell et al. (1997), Mikalsen and Jentoft (2001), Halim et al. (2008) and others add legal/ethical legitimacy as a third dimension of stakeholder identification, but this is to name the problem rather than solve it. Furthermore, ethical and legal legitimacy may come apart, creating further difficulties.

The question of legitimacy is often ignored or placed in the background by authors in the social science and management literatures who discuss stakeholder

[76] Laura Dunn's 2007 documentary *The Unforeseen* describes just such a conflict surrounding proposed development around Barton Springs in Austin, Texas, a local landmark and home to the endemic Barton Springs Salamander (*Eurycea sosorum*).

approaches. Grimble and Wellard (1997) provide a review of stakeholder methodology in the related field of natural resource management. Stakeholders are distinguished in terms of a continuum of spatial and temporal scale: “micro” stakeholders range from current local to regional interests while “macro” stakeholders range from national and international interests to the interests of future generations. Conflicts within a single stakeholder’s goals (termed tradeoffs) are distinguished from conflicts between the diverse stakeholders and stakeholder groups (termed conflicts). Despite the interest and usefulness of these particular distinctions, the normative question of legitimacy is avoided altogether.

More recently, Reed et al. (2009) reviewed methods of stakeholder identification, stakeholder categorization, and the investigation of relationships between stakeholders. However, the focus was on strengths and weaknesses of various social scientific methodologies, for example, the use of focus groups, structured interviews, or snowball sampling to identify stakeholders. Little to no attention was paid to normative issues, although Reed et al. do note, for example, that some methods are more likely than others to exclude marginalized groups. Specifically, the use of influence-interest matrices, similar to the influence-urgency diagram in Figure 1, which categorize stakeholders according to how much they are influenced by a decision and how much they are interested in the outcome, may lead to biased consultation with more influential groups. However, how or why marginalized groups should be included, or whether interest or influence alone could make one a legitimate stakeholder, or whether a combination of interest and influence are required, are not discussed.

Stakeholder methodologies often treat each interest group as roughly equal, an assumption challenged by Singleton (2009) in her examination of the role of indigenous groups in debates over marine protected areas in Washington State (U.S.A). Singleton argues that the role of political institutions in generating asymmetries of power and influence is often obscured by stakeholder methodologies that make this equality assumption. She stresses that the establishment of marine protected areas has important effects on the allocation of fisheries resources, as tradeoffs benefit some stakeholder groups at the expense of others.

The particular legal and historical context is especially salient here: a U.S. federal court ruled in 1974 (in *United States vs. Washington*) that treaties from the mid-19th century entitled the 12 signatory tribes to 50% of Washington's harvestable salmon stocks in their "usual and accustomed" fishing areas; this ruling was extended to shellfish in the 1990s. The ruling has allowed these tribes to develop significant power in Washington's politics of natural resource management. However the strained historical relationship between the U.S. government and Washington tribal groups still affects negotiations over marine protected areas, where fishing is legally limited.

Tribal fishermen harbor resentment towards conservationists due to past abuses by the government: according to Singleton, "[in] the decades leading up to the U.S. vs. Washington case, tribal fishermen were prosecuted, jailed, and their equipment seized or destroyed—ostensibly on the grounds of conservation" (Singleton, 2009, 426). Tribes are especially concerned that the benefits and burdens of marine protected areas will not be shared equitably. They claim that marine conservation proponents are using conservation as a pretext for what is actually resource allocation, since tribal fishing rights are place-specific, whereas non-treaty fishing rights are not. The argument is thus that the burdens of conservation are not distributed equitably.

One might argue that the tribes deserve special status as stakeholders due to the historical injustices they have faced, from which many still suffer. As Singleton points out, it may *appear* fair to limit *all* fishing (tribal and non-tribal) in marine protected areas, but tribal fishermen are arguably affected more than others, due to place-specificity of their fishing rights and the special role fishing plays in tribal life. Furthermore Washington state has allowed economic activity that indirectly affects salmon stocks, for example by allowing water to be diverted for hydroelectric power generation, without tribal approval. On the other hand, one might argue that the tribes' current legal and political power is adequate compensation for historical injustice, and so the tribes should be treated just as any other stakeholder in negotiations surrounding marine protected areas. Either way, this case illustrates that power asymmetries and historical context complicate the stakeholder approach, which on its own cannot deal with questions of social justice and the equitable distribution of benefits and burdens.

### 5.2.2. Moral Status, Stakeholders, and Rights Holders

Another problem with the stakeholder methodologies is that they seem to rule out certain groups, particularly non-human animals and future generations, who many regard as having moral status. What is usually meant by the question of moral status is whether someone or something's interests should "count" morally. Environmental ethicists who advocate for the moral status of non-humans, for example sentientists[77] like Singer (1975), would object to the implicit anthropocentrism of the stakeholder literature, which always assumes stakeholders are humans. If a land use decision would result in a large population of sentient animals being killed or driven from their habitat, the sentientist would argue that it would be morally wrong to ignore this consideration, and would probably place these animals in the "high urgency, low power" stakeholder category in Figure 1.

---

[77] A sentientist believes that sentience, or the capacity to feel pleasure or pain, is sufficient for moral status. Another way of putting this is that all pleasure and pain interests should be taken into account in moral decision making.

There are practical reasons for the restriction to humans, since stakeholder-oriented approaches in biological conservation often seek to identify stakeholders in order to include them in processes of planning and decision-making (Reed 2008), or else simply to identify their interests in order to be prepared to deal with potential conflicts. For obvious reasons it is difficult to include non-human animals, or future generations of humans for that matter, in processes of decision-making, and these groups have little or no power to resist decisions once made.

Furthermore, someone may have moral status in the philosopher's sense, but not be an ethically legitimate stakeholder for a particular land use decision, since we would not accept that their relevant interests should be taken into account at all in that particular decision. The simplest examples here are humans with unsavory preferences, for example, people interested in killing individuals belonging to an endangered species for sport. While conservationists would be inclined to ignore this particular preference in a land use decision involving the endangered species for ethical reasons, this does not imply that the hunters lack moral status.

This suggests two distinctions. Firstly, a distinction between the question of moral status "full stop," whether someone or something's interests is morally considerable and thus should be taken into account by decision-makers at all, and more complex evaluative questions of moral status, which concern how one ought to weigh these various claims and interests. For any particular case, how to weigh the interests of "micro" versus "macro" stakeholders, for example the interests of current local residents versus the interests of future generations, just is the relevant normative question. For example, environmental ethicists advocating environmental justice often focus on cases where the relevant agents are all assumed to be legitimate stakeholders to some degree, but there are unjustified asymmetries in decision-making power and/or environmental outcomes.[78]

Secondly, ethicists would draw a further distinction between stakeholders as merely interested parties and rights holders, whose rights claims, where legitimate, should be respected even when a violation of that right could be traded off against significant gains to other interests (Dworkin 1984, Wenar 2011). For example, perhaps the interests of vulnerable populations of traditional resource users with unique cultures and legitimate claims to land should in some cases trump the interests of land-development capital or conservation scientists (Dowie 2009). Such rights could be justified on a social contractarian basis (Rawls 2001), as following from rules we would all agree upon in a fair decision procedure, or else on a consequentialist basis (Kagan 1997), as their respect may lead to better consequences than situations in which they are not respected.

Whether such purported rights are absolute, implying that their violation should *never* be traded off against any gains, or they admit of lexicographic threshold effects such that *some* amount of gains could be traded off against the violation of the right, they should be distinguished from fully commensurable values which may be traded off in the manner of MAVT. However, deciding in practice whether such rights exist and how they should be weighed against other interests and considerations constitute significant challenges. The next section discusses tradeoffs in conservation planning and a case study of difficult tradeoffs, focusing on a case where animal welfare claims must be weighed against multiple and competing biological values.

[78] See, for example, Schrader-Frechette (2002), Estrella-Luna (2010), and Nadasdy (2005).

## 5.3. Weighing Multiple Values: Ethics and Difficult Tradeoffs in Biological Conservation

### 5.3.1. Tradeoffs in Conservation Planning

Biodiversity conservation planning, for example within the Systematic Conservation Planning protocol (Margules and Sarkar 2007), necessitates the identification of biodiversity constituents and/or surrogates. Biodiversity constituents are the aspects of the biota that are deemed worth preserving (they may be species but need not be), while surrogates are putatively correlated measures of those constituents used for planning purposes. As discussed in chapter 1, the identification of biodiversity constituents is primarily a normative question about society's values, and will necessarily imply tradeoffs, since resources are scarce, for conservation and otherwise.

While tradeoffs between aspects of the non-human biota at various scales are inevitable when deciding what to actively conserve or manage, perhaps the most ethically interesting and problematic tradeoffs are those between biological or ecological units and human welfare interests (McShane et al. 2010), where concerns for social justice meet environmental ethics (Brechin et al. 2002). The best examples of conservation planning exercises, where areas are prioritized for conservation management as opposed to extractive use, attempt to explicitly account for these tradeoffs.

For example, consider the conservation plan for Papua New Guinea (PNG) proposed by Faith et al. (2001). Planners prioritized areas by incorporating data on other land uses, particularly agriculture and forestry, as well as population density and previously conserved areas. Opportunity costs were integrated into the selection process by using indices of timber volume and agricultural potential: the PNG Forest Authority provided planners data on timber volume, and the PNG Department of Agriculture and Livestock proposed a simple model of agricultural potential based on slope and drainage classes. In the selection process, preference was given to units having low agricultural potential and timber volume, and planners “masked out” or excluded units with high land use intensity. Conservation Needs Assessment high priority areas[79] were also given preference, as were areas with low human population density, while previously conserved areas were automatically incorporated into the final plan. Planners computed the most economical set of planning units using these multiple criteria.

The PNG planners incorporated data on opportunity costs and tradeoffs at the beginning of the process. Ethical reasons to consider such tradeoffs stem from concern for the humans who live there. 37% of PNG’s extremely culturally diverse population lives below the poverty line, and 85% of the population make a living by subsistence agriculture (CIA 2009). Thus a refusal to select biological conservation priority areas by incorporating economic, especially agricultural, tradeoffs would arguably be unethical.

While it is easy to say that the mere consideration of tradeoffs is an ethical imperative in the presence of scarce resources, actually weighing and trading off multiple values in an ethically acceptable way is more difficult. The following case study deals with introduced North American beavers in Navarino Island, Chile. Here several questions arise. The first is one of identifying valuable biodiversity and trading off different biological units: should the beavers be considered a threat to “native” biodiversity, or a potential asset? Furthermore, how should the beavers’ welfare rights weigh (if at all) in determining a management strategy? Finally, how should the values of conservationists, many of whom advocate total extirpation of the beavers, be weighed against local people who may have different values?

---

[79] These areas were identified by conservation biologists commissioned by the government of PNG in the early 1990s, and included “sites of high endemism, high species richness, and unusual ecosystems and habitats” (Swartzendruber 1993, ix). Thus an argument could be made that these conservation biologists, while they did take into account economic tradeoffs, did not take into account the values of the people of PNG vis-à-vis biological priorities.

### 5.3.2. Beavers and Ecological Tradeoffs in Navarino Island, Chile

#### 5.3.2.1. Background and Multiple Values

Consider the case of the introduced North American beaver (*Castor canadenis*) on Navarino Island in southern Chile. The beaver, whose original range spanned from northern Mexico through most of North America, was introduced to Tierra del Fuego Island in 1946 and quickly became established there, the mainland, and Navarino, Picton, Lennox, Nueva, Dawson, and Hoste islands. According to Skews et al. (1999), approximately 20,000 beavers occupied Navarino Island in the year 2000, at a density of 1.1 colonies per $km^2$. Worries about the significant ecological effects of beavers' felling trees and constructing dams have motivated concern among some conservationists and residents, while others see the beavers as a potential environmental and economic asset.

Alternative management plans under consideration include eradication of the entire beaver population, favored by the Argentinean and Chilean governments and many conservationists, which would constitute the largest official eradication program by land area in history (Choi 2008). Alternatives short of complete extirpation include some degree of control of the beaver population, perhaps with the establishment of some no hunting zones to protect a small population of beavers (Schüttler et al. 2011). This decision raises normative questions about the value of the beavers and the habitat changes they cause, as well as questions of how we ought to treat or humanely kill the beavers, which like all rodents are generally accepted to be sentient.

Haider and Jax (2007) summarize many arguments from anthropocentric and non-anthropocentric points of view, in terms of material costs and benefits and non-material values like beauty, rarity, and species richness. They cite the fact that the beavers have been known to occasionally cause damage to roads, irrigation channels, sewers, and other infrastructure, as well as soil erosion and increased sedimentation by felling trees. On the other hand, beaver fur, meat, and secretion products have potential economic and cultural value, and apparently the beaver has become something of a mascot for some local residents.[80] The effects of the beaver on economic benefits from tourism are less clear. Beaver dams impede hiking, but may attract some tourists interested in the beaver's role as ecosystem engineer. Aesthetic appreciation of the beaver's engineered habitat may have value to some, but others, including the Chilean government's Agricultural and Livestock Service, see the beavers as invasive pests.

In their qualitative interview study of local attitudes toward the beaver and other populations of non-native species on Navarino, Schüttler et al. (2011) discovered that local people associated the beaver with several categories of positive values, including (using the classification in Kellert 1996) utilitarian (consumptive), naturalistic (satisfaction through experiences with beavers), ecological-scientific, aesthetic, symbolic (particularly as a "mascot"), and humanistic (emotional attachment and identification with beavers, and keeping them as pets). However they were also identified as pests by some interviewees, while others reported being disgusted by them.

Ecologically, beaver dams and their resulting riparian zones may be leading to the elimination of southern beech (genus Nothofagus) populations (Pastur et al. 2006). However the beavers' activities have resulted in increased habitat and breeding ground for exotic trout species (*Oncorhynchus mykiss* and *Salmo trutta*) as well as some waterfowl species (e.g. Ashy-headed goose [*Chloephaga poliocephala*] and Yellow-billed teal [*Anas flavirostris*]). Furthermore, Wright et al. (2002) argue that the beaver's role as ecosystem engineer can increase plant species richness at the scale of an entire

[80] In their interview study of local perceptions of the beaver Schüttler et al. (2011, 179) quote one resident as saying, "It is like our mascot," and a member of the Navy as saying "…you see a beaver and suddenly you feel happy…especially in winter times when there is an ice cap, and you see them swimming underneath."

landscape (i.e. an area with several patch types) by enhancing habitat heterogeneity. Abandoned beaver ponds are often succeeded by meadows, precipitated by dams trapping nutrient-rich sediment. So beavers might be taken as a surrogate for plant biodiversity, although a surrogacy analysis has not been performed on these populations. Either way, evaluating the outcomes where the beavers are left alone, their population is significantly controlled, or they are extirpated, will necessitate considering these ecological tradeoffs.

#### 5.3.2.2. Ethical Aspects

From an ethical point of view, the case of *Castor canadenis* in southern Chile is not as clear-cut as it would seem listening to passionate opponents of invasive species, like those quoted in Choi (2008).[81] The beavers' recent introduction, and "invasive" ability to reproduce and disperse throughout the archipelago in the absence of predation, is not a sufficient reason to call for their extirpation, since, at a certain scale, this argument would equally apply to *Homo sapiens*. However the ability of the beaver to quickly reproduce and disperse is an important instrumental concern that has implications for the costs of extirpation or management, as complete extirpation may not actually be feasible.

It is easy to say that an effective argument for extirpation would have to show that the harm beavers cause to the human population and native biodiversity outweighs the significant financial, logistic, and ethical costs associated with this plan. Even if the beavers' activity causes decline in populations of certain native species, it remains an open question whether these native species are worth saving at the cost of removing the beavers. Indeed, some may welcome landscape change and the resulting shift in local

[81] For example, Choi quotes one American ecologist saying that the destruction to the forest caused by the beavers is "like bulldozers steamed through."

biodiversity, advocating control of the beaver population falling short of complete eradication.

According to Schüttler et al. (2011, 180), moderate population control is the majority position amongst local people interviewed, who were concerned with the costs, feasibility, and ethicality of extirpation. Nature conservationists were apparently an exception, most of whom favored extirpation. A game theoretical analysis of the situation on Navarino could illuminate potential strategic interactions between these groups, especially if some individuals concerned with the beavers were willing to act to save some, thwarting attempts at complete extirpation. Say that such an analysis identified a situation structured like the ones discussed in chapter 4, where individuals acting rationally according to their own interest cause a situation that is worse for all agents. For example, consider a situation in which conservationists and the government invest significant resources in a massive extirpation project, but the project fails due to the actions of a few local people. This outcome might be preferred less by all agents than a situation in which local people exert moderate local control over beaver populations, leaving larger populations of beavers but costing much less. However, from an ethical point of view, this would not constitute a successful normative argument for cooperation between these groups *on its own*, since this would require the assumption that all agents' preferences are reasonable, and the cooperative outcome is ethically realized.

Additionally, while MAVT would be useful for explicitly considering these tradeoffs quantitatively, in the absence of agreement amongst stakeholders over relevant criteria of evaluation, their weights, and method of aggregation, these models can provide little guidance. Schüttler et al. (2011, 182) recommend collaborative stakeholder workshops in the absence of shared values; however as argued above, identification of the relevant stakeholders also raises questions about the legitimate roles of local and national stakeholders in making decisions, and how participation of stakeholders should take place in negotiation with scientific expertise is an ongoing and controversial area of research.[82]

Ethical problems associated with eradicating the beavers raise further difficulties for an analysis of tradeoffs. Any extirpation plan would obviously involve causing suffering and death to a large population of sentient mammals. The beavers are usually trapped using steel conibear traps, which are designed to kill animals quickly by snapping down on the neck or head, usually closing the trachea and/or fracturing the spine. However these traps can also accidentally kill other animals like large birds or dogs.

While sentientists would likely be willing to trade off pleasures to humans and pains to beavers (and perhaps would also be willing to trade off the pain of the beavers for the "higher" pleasures enjoyed by humans who enjoy native biological diversity), they would not accept that the beavers could be killed merely for the benefit of non-sentient native trees like the southern beech. Stronger advocates of animal rights would perhaps contend that the right of the beavers not to be killed strongly trumps the removal of inconveniences to humans. How these concerns for animal welfare and/or rights should be weighed against the damage the beavers cause to the ecosystem and human infrastructure, for example, is an ethical question that cannot be resolved through decision analytic techniques alone, whose purpose is to elicit or construct preferences whose content may or may not lead to ethical decisions.

This case illustrates the point that identifying biodiversity constituents and determining a management strategy involves trading off parts of the biota. If the ecosystem engineering beavers are removed, the fish and bird species that benefit from

[82] See, for example, Chess and Purcell (1999), Beirle and Konisky (2001), and Reed (2008).

beaver habitat will also decline, whereas southern beech and other populations will benefit. The beavers will no longer cause damage to human infrastructure, but the costs of extirpation are also significant. The question of how to manage beaver populations in southern Chile cannot be resolved by simply considering a single dimension of value, and the use of formal tools like MAVT for the analysis of tradeoffs would have to weigh the rights and interests of numerous stakeholders and rights holders, potentially including humans and non-humans. Game theoretical analyses like those of chapter 4 may be useful for revealing strategic interactions between stakeholder groups, but the argument from cooperation implicitly assumes that the preferences of the relevant agents are reasonable and the cooperative outcome is not only Pareto-efficient, but better all things considered.

## 5.3. Concluding Discussion: The Value and Limits of Decision Science

As long as we accept *some* justification for biological conservation, then trading off other goods for conservation is unavoidable in a world of scarce resources. Furthermore, as the example from conservation planning in PNG illustrates, explicit consideration of these difficult tradeoffs and conflicts is ethically preferable to obscuring or ignoring them. However, biological conservation contexts reveal tradeoffs that are particularly troubling ethically: we are forced to weigh goods as distinct as biodiversity and social justice, or biodiversity and human or non-human animal welfare.

Problems with his overall position notwithstanding,[83] Rolston's (1996) infamous claim that nature reserves should sometimes be prioritized over rescuing starving humans just follows from a point of view that does not see saving starving humans, despite its obvious ethical importance, as lexicographically more preferred to all other pursuits.

[83] For apt critiques, see especially Guha (1989) and Siurua (2006).

Anyone who believes it is morally acceptable to donate money to a university or an art museum when that money could go to save a starving child implicitly accepts this point of view. As Rolston puts it (2006, 265), “Human rights to development, even by those who are poor, though they are to be taken quite seriously, are not everywhere absolute, but have to be weighed against the other values at stake.”

The main strength of the formal tools of decision science explored in this dissertation is that they allow decision-makers to make the multiple values at stake in biological conservation decisions more explicit. The problem identified in chapter 1 was that there are multiple ways ‘biodiversity’ can be construed and biological diversity measured, and which definition or measure used in applied context will depend on the values of the investigators. However, such values are often implicit and thus obscure. The resulting transparency from applying decision theory allows decision-makers to think about the logical consequences of their values and the values of other agents, whether by considering tradeoffs as in chapters 2 and 3 or in interdependent decisions as in chapter 4. This transparency also facilitates reflection, as decision-makers are forced to confront difficult conflicts of values by constructing preferences over tradeoffs, and multiple agents may realize that while their individual strategies seem rational, they will lead to outcomes that are worse for everyone.

Of course, while decision theory places logical constraints on the structure of values (for example transitivity of preferences), it does not place any constraints on their content. While some attempts have been made to add axioms to standard decision theory to accommodate the structure of particular ethical theories,[84] the resulting axiom systems simply add more definitions and structural constraints, for example by defining utilitarian

[84] For a recent attempt, see Colyvan et al. (2010).

social welfare functions as an equally weighted sum of individual utility functions.[85] Thus for the argument from commensurability or the argument from cooperation to be normative, further ethically substantive assumptions must hold. Construed as normative ethical arguments, the argument from commensurability and the argument from cooperation are both consequentialist (Kagan 1997), although the latter has social-contractarian aspects (Gauthier 1986).[86] The argument from commensurability recommends weighing multiple values and choosing an alternative whose consequences achieve the best balance of these values. The argument from cooperation recommends negotiating an agreement or social contract to achieve an outcome that has better consequences for all parties concerned than the outcome in the absence of such an agreement. Both types of argument, to have normative force, rely on several further assumptions of ethical reasonableness, enumerated in the next paragraphs. Here, by '*x* is ethically reasonable' I just mean that *x* would not be ruled out by ethical reflection, potentially including but not requiring input from ethical theories.

The argument from commensurability assumes that all relevant values can be specified, and those values are ethically reasonable. It also assumes that those values can be placed on a common scale, weighted, and aggregated, and that the judgments on weights and aggregation represent ethically reasonable judgments about tradeoffs between the multiple values. It further assumes that rights claims and other lexicographic priorities can be respected, perhaps as constraints on the alternatives considered, or else are absent.

---

[85] Normative questions raised by defining such a function include: Whose utility functions? How should they be calibrated? Should they be based on subjective welfare or an objective list of goods that satisfy basic human needs?

[86] Whether the latter is best construed as consequentialist or contractarian depends on the justification for the social contract, whether good consequences (consequentialist) or some form of consent or as the outcome of a fair decision procedure (more along the lines of traditional social contractarianism).

The argument from cooperation assumes that all agents relevant to an interdependent decision can be specified, and their values are ethically reasonable. In the original formulation of the prisoner's dilemma, two criminals are trying to escape jail time. If both cooperate, they will escape jail, but each has an individual incentive to defect. While the argument from cooperation shows that a cooperative social contract would be instrumentally rational for the prisoners, society would rightly object on ethical grounds: the argument is normative only for agents with ethically reasonable preferences. Furthermore the argument assumes that agents will be able to formulate a social contract in an ethically reasonable way, either through intervention by a third party, or through repeated interactions and reciprocity.

As mentioned in the introduction, models from decision science cannot bypass informal deliberation or ethical reflection. These are required for inputs to the models to be ethically reasonable and thus for the arguments to be normative. However it does offer tools that allow rigorous ethical arguments to be made once the required assumptions are met.

# Chapter 6: Summary and Conclusions

The arguments of this dissertation have largely been concerned with the scope and limits of normative applications of models from the decision sciences in complex decisions in biological conservation characterized by multiple values and multiple agents. The first chapter motivated the use of multi-criteria approaches to decision-making in biological conservation by analyzing the concept of biodiversity, arguing that while it is often taken as a general goal of biological conservation, its meaning is often obscure, especially since its most catholic usage refers to "the variety within and among living organisms, assemblages of living organisms, biotic communities, and biotic processes" (DeLong 1996). Several things could be meant by 'biodiversity' and biological diversity can be measured at many scales, so the determination of its meaning in applied scientific contexts (for example, as measured by species richness) where conservation goals are pursued carries risks that some relevant biological values (for example, endemism) will be excluded. Here I also suggested that the value of biodiversity is dependent on facts about biological composition, calling into question the idea that maintaining or conserving biodiversity *per se* should be the main goal of conservation biology as opposed to one of many goals.

The solution canvassed in chapter 2 is to make multiple values explicit and to use techniques from decision analysis, particularly multi-attribute value theory, to weigh these multiple values and to allow for quantitative consideration of complex tradeoffs. Here I argue that constructing preferences over complex tradeoffs ("constructing commensurability") can be a practical imperative due to the existence of psychological tendencies to avoid tradeoff thinking and use simple heuristics with non-normative implications. However, technical and ethical limitations constrain the use of such approaches. Chapter 3 contributed a case study of constructing commensurability, where I examined the system used by the Fish and Wildlife Service to rank National Wildlife Refuges for budgeting purposes. Several flaws of this system were identified, including potential failures of the assumptions discussed in the previous chapter that are required for the use of additive value functions.

A normative role for game theory is identified in chapter 4, which discusses biological conservation decisions where multiple agents interact. While formally isomorphic to decision problems with multiple values, in practice these problems are very different, since multiple agents determine an outcome by autonomously acting in their own interests. Here I identify cases where decision-makers acting in their own interests cause an outcome that is worse for everyone relative to a situation in which they coordinate their behavior. I argue that while multiple solutions to such problems of cooperation exist, solutions should take into account results from behavioral game theory, which show that people are often willing to cooperate in reciprocal relations over time and enforce norms of cooperation at their own expense.

The final chapter discusses ethical dilemmas and normative problems for analyzing biological conservation decisions, arguing that the models from decision science discussed earlier cannot work without input from normative and applied ethics. Here I enumerate ethical assumptions that must hold for the arguments from commensurability and cooperation to have normative force, particularly that the input valuations should be ethically reasonable, and identify problems for stakeholder methodologies in identifying and weighing the interests of multiple agents.

The main conclusions of this dissertation are:

1. The multiplicity of meanings of 'biodiversity' and measures of biological diversity raise risks for conservation biology and motivate multi-criteria approaches to conservation decision-making;
2. Constructing commensurability between multiple values to explicitly and quantitatively analyze tradeoffs is a practical imperative in biological conservation decisions with high stakes and complex tradeoffs, but applications should be wary of the limitations of such techniques, particularly their strong ethical and technical assumptions;
3. Game theory can play a normative role in biological conservation decisions by identifying situations where individuals acting independently in their own interest cause an outcome that is worse for everyone relative to cooperative outcomes, but potential solutions to such dilemmas should not ignore empirically established human dispositions to enforce norms and cooperate in repeated interactions;
4. Decision science can aid in making values explicit, facilitating reflection and learning, but cannot resolve ethical dilemmas on its own without input from normative and applied ethics, particularly in identifying legitimate stakeholders and weighing multiple biological concerns against concerns for rights, welfare, and social justice. The arguments from commensurability and cooperation require substantive ethical assumptions to have normative force.

## References


Aldred, J. 2006. "Incommensurability and Monetary Valuation." *Land Economics*, 82, 2, 141-161.

Anand, P. 1993. "The Philosophy of Intransitive Preferences." *The Economic Journal*, 103, 337-346.

Arrow, K. 1951. *Social Choice and Individual Values*. New Haven, CT: Yale University Press.

Arrow, K. and H. Reynaud. 1986. *Social Choice and Multicriterion Decision-making*. Cambridge, MA: MIT Press.

Anderson, E. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Beattie, J. and S. Barlas. 2001. "Predicting Perceived Differences in Tradeoff Difficulty." In E.U. Weber, J. Baron, and G. Loomes eds. *Conflict and Tradeoffs in Decision Making*. New York, NY: Cambridge University Press, 25-65.

Beirle, T.C. and D.M. Konisky. 2001. "What are We Gaining from Stakeholder Involvement? Observations from Environmental Planning in the Great Lakes." *Environment and Planning C: Government and Policy*, 19, 515-527.

Beissinger, S.R. and D.R. McCollough. 2002. *Population Viability Analysis*. Chicago, IL: University of Chicago Press.

Beissinger, S.R. and M.I. Westphal. 1998. "On the Use of Demographic Models of Population Viability in Endangered Species Management." *The Journal of Wildlife Management*, 62, 3, 821-841.

Bermúdez, J.L. 2009. *Decision Theory and Rationality*. New York, NY: Oxford University Press.

Binmore, K. 2008. *Rational Decisions*. Princeton, NJ: Princeton University Press.

Bowles, S. 2008. "Policies Designed for Self–interested Citizens May Undermine 'the Moral Sentiments': Evidence from Economic Experiments." *Science*, 320, 1605-1609.

Bowles, S. and S. Hwang. 2008. "Social Preferences and Public Economics: Mechanism Design when Preferences Depend on Incentives." *Journal of Public Economics*, 92, 1811-1820.

Boyce, M.S. 1992. "Population Viability Analysis." *Annual Review of Ecology and Systematics*, 23, 481-506.

Brans, J.P. and P. Vincke. 1985. "A Preference Ranking Organization Method: The PROMETHEE Method for Multiple Criteria Decision-Making." *Management Science*, 31, 6, 647-656.

Brechin, S.R., Wilshusen, P.R., Fortwangler, C.L. and P.C. West eds. 2002. *Contested Nature: Promoting International Biodiversity with Social Justice in the Twenty-first Century*. Albany, NY: SUNY Press.

Broome, J. 1997. "Is Incommensurability Vagueness?" in R. Chang ed. 1997, 67-89.

Butler, J.C., Dyer, J.S., and J. Jia. 2006. "Using Attributes to Predict Objectives in Preference Models." *Decision Analysis*, 3, 2, 100-116.

Camerer, C.F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
Cardenas, J.C., Stranlund, J., Willis, C. 2000. "Local Environmental Control and Institutional Crowding-out." *World Development*, 28, 1719-1733.
Carlson, E. 2001. "Organic Unities, Non-Trade-Off, and the Additivity of Intrinsic Value." *The Journal of Ethics*, 5, 335-360.
Carson, R. 2011. *Contingent Valuation: A Comprehensive Bibliography and History*. Cheltenham: Edward Elgar Publishing.
Chang, R. ed. 1997. *Incommensurability, Incomparability, and Practical Reason*. Cambridge, MA: Cambridge University Press.
Chang, R. 2001. "Against Constitutive Incommensurability or Buying and Selling Friends." *Philosophical Issues*, 11, 33-60.
Chapin, M. 2004. "A Challenge to Conservationists." *WorldWatch* November/December, 17-31.
Chess, C. and K. Purcell 1999. "Public Participation and the Environment: Do We Know What Works?" *Environmental Science and Technology*, 33, 16, 2685-2692.
Choi, C. 2008. "Tierra del Fuego: the Beavers Must Die." *Nature*, 453, 968.
CIA. 2009. *The World Factbook 2009*, Central Intelligence Agency. URL = <https://www.cia.gov/library/publications/the-world-factbook/index.html>. Accessed November 2011.
Clemen, R.T. and J.E. Smith. 2009. "On the Choice of Baselines in Multiattribute Portfolio Analysis: A Cautionary Note." *Decision Analysis*, 6, 4, 256-262.
Colyvan, M., Cox, D., and K. Steele. 2010. "Modeling the Moral Dimension of Decisions." *Nous*, 44, 3, 503-529.
Creel, S., and N. Creel. 2002. *The African Wild Dog: Behavior, Ecology, and Conservation*. Princeton, NJ: Princeton University Press.
Dancy, J. 2004. *Ethics Without Principles*. New York, NY: Oxford University Press.
Dancy, J. 2009. "Moral Particularism." In E. Zalta ed. *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/spr2009/entries/moral-particularism/>. Accessed November 2011.
Deci E., Koestner R., and R. Ryan. 1999. "A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin,* 125, 627-688.
DeLong, D.C. 1996. "Defining Biodiversity." *Wildlife Society Bulletin,* 24, 738-749.
Donaldson, T., Preston, L., 1995. The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications. *The Academy of Management Review,* 20, 1, 65-91.
Douglas, H. 2000. "Inductive Risk and Values in Science." *Philosophy of Science,* 67, 4, 559-579.
Douglas, H. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh, PA: University of Pittsburgh Press.
Dowie, M., 2009. *Conservation Refugees: The Hundred-Year Conflict between Global Conservation and Native Peoples*. Cambridge, MA: MIT Press.
Dworkin, R. 1984. "Rights as Trumps." In J. Waldron ed. *Theories of Rights*. Oxford:

Oxford University Press. 153-167.
Dyer, J.S. and R.K. Sarin. 1979. "Measurable Multiattribute Value Functions." *Operations Research*, 27, 4, 810-822.
Edwards, W. and D. von Winterfeldt. 1986. *Decision Analysis and Behavioral Research*. New York, NY: Cambridge University Press.
Elliott, K. 2009. "The Ethical Significance of Language in the Environmental Sciences: Case Studies from Pollution Research." *Ethics, Place, and Environment*, 12, 2, 157-173.
Ellis, S. 2008. "The Main Argument for Value Incommensurability (and Why It Fails)." *The Southern Journal of Philosophy*, 46, 27-43.
Estrella-Luna, N. 2010. "Public Participation and Communicative Interaction: The Structural Mechanisms of Institutional Bias." *Environmental Justice*, 3, 4, 135-140.
Faith, D. 1992. "Conservation Evaluation and Phylogenetic Diversity." *Biological Conservation,* 61, 1-10.
Faith, D., Margules, C., and P. Walker. 2001. "A Biodiversity Conservation Plan for Papua New Guinea based on Biodiversity Trade-offs Analysis." *Pacific Conservation Biology,* 6, 304-324.
Falkinger, J., Fehr, E., Gaechter, S., and R. Winter-Ebmer. 2000. "A Simple Mechanism for the Efficient Provision of Public Goods—Experimental Evidence." *American Economic Review*, 90, 247-264.
Fehr, E. and U. Fischbacher. 2004. "Third-party Punishment and Social Norms." *Evolution and Human Behavior*, 25, 63-87.
Fehr, E., Fischbacher, U. and S. Gächter 2002. "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms." *Human Nature*, 13, 1-25.
Fehr, E. and S. Gächter 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90, 980-994.
Fehr, E. and B. Rockenbach. 2003. "Detrimental Effects of Sanctions on Human Altruism." *Nature*, 422, 137-140.
Fieberg, J. and S.P. Ellner. 2000. "When is it Meaningful to Estimate an Extinction Probability?" *Ecology*, 81, 7, 2040-2047.
Fiorino, D. 1990. "Citizen Participation and Environmental Risk: a Survey of Institutional Mechanisms." *Science, Technology, and Human Values*, 15, 2, 226-243.
Fischoff, B. 1991. "Value Elicitation: Is There Anything in There?" *American Psychologist*, 46, 8, 835-847.
Fischoff, B. 2006. "Constructing Preferences from Labile Values." In Lichtenstein and Slovic, eds. 2006, 653-668.
Fischoff, B., Slovic, P., and S. Lichtenstein. 1980. "Knowing What you Want: Measuring Labile Values." In T. Wallstein, ed. *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, NJ: Erlbaum, 117-141.
Fiske, A. and P. Tetlock. 1997. "Taboo Tradeoffs." *Political Psychology*, 18, 2, 255-297.
Frank, D.M. and S. Sarkar. 2010. "Group Decisions in Biodiversity Conservation:

Implications from Game Theory." *PLoS ONE*, 5, 5.
Freeman, R., 1984. *Strategic Management: A Stakeholder Approach*. Boston, MA: Pitman.
Frey, B.S. 1997. "A Constitution for Knaves Crowds out Civic Virtues." *The Economic Journal*, 107, 443, 1043-1053.
Gauthier, D. 1986. *Morals by Agreement*. New York, NY: Oxford University Press.
Gaston, K. 1996. *Biodiversity: A Biology of Numbers and Difference*. Oxford: Blackwell.
Gerber, N. 2011. "Biodiversity Measures Based on Species-level Dissimilarities: A Methodology for Assessment." *Ecological Economics,* 70, 2275-2281.
Georgescu-Roegen, N. 1954. "Choices, Expectations, and Measurability." *Quarterly Journal of Economics*, 68, 4, 503-534.
Gintis, H. 2000. "Beyond Homo Economicus: Evidence from Experimental Economics." *Ecological Economics*, 35, 311–322.
Gintis, H. 2009. *Game Theory Evolving*, 2nd ed. Princeton, NJ: Princeton University Press.
Glimcher, P. 2011. *Foundations of Neuroeconomic Analysis*. New York, NY: Oxford University Press.
Glowka, L., Burhenne-Guilmin, F., Synge, H., McNeely, J. A. McNeely and L. Gundling. 1996. "A Guide to the Convention on Biological Diversity." *Environmental Policy and Law Paper No. 30*. Gland, Switzerland: IUCN.
Gneezy U. and A. Rustichini. 2000. "A Fine is a Price." *The Journal of Legal Studies,* 29, 1–17.
Golabi, K., Kirkwood, C.W., and A. Sicherman. 1981. "Selecting a Portfolio of Solar Energy Projects Using Multiattribute Preference Theory." *Management Science*, 27, 2, 174-189.
Gordon, H.S. 1953. "An Economic Approach to the Optimum Utilization of Fisheries Resources." *Journal of the Fisheries Research Board of Canada*. 10, 442-457.
Gordon, H.S. 1954. "Economic Theory of a Common Property Resource: the Fishery." *Journal of Political Economy*, 62, 124–142.
Griffin, J. 1997. "Incommensurability: What's the Problem?" in R. Chang, ed. 1997. 35-51.
Grimble, R. and K. Wellard. 1997. "Stakeholder Methodologies in Natural Resource Management: a Review of Principles, Contexts, Experiences and Opportunities." *Agricultural Systems,* 55, 2, 173-193.
Guha, R. 1989. "Radical American Environmentalism and Wilderness Preservation: a Third World Critique." *Environmental Ethics*, 11, 71-83.
Gusset, M., Maddock, A.H., Gunther, G.J., Szykman, M., Slotow, R., Walters, M., and M.J. Somers. 2008. "Conflicting Human Interests Over the Re-introduction of Endangered Wild Dogs in South Africa." *Biodiversity and Conservation,* 17, 83-101.
Haider, S. and K. Jax. 2007. "The Application of Environmental Ethics in Biological

Conservation: a Case Study from the Southernmost Tip of the Americas." *Biodiversity Conservation,* 16, 2559–2573.
Halim, A., Soekirman, T., and W. Ramono. 2008. "Involving Resource Users in the Regulation of Access to Resources for the Protection of Ecosystem Services Provided by Protected Areas in Indonesia." In N. Sodhi, G. Acciaioli, M. Erb, and K. Tan eds., *Biodiversity and Human Livelihoods in Protected Areas: Case Studies from the Malay Archipelago*. New York, NY: Cambridge University Press. 122-138.
Hardin, G. 1968. "Tragedy of the Commons." *Science*, 162, 1243-1248.
Hausman, D. 2000. "Revealed Preference, Belief, and Game Theory." *Economics and Philosophy*, 16, 99-115.
Heidenberger, K. and C. Stummer. 1999. "Research and Development Project Selection and Resource Allocation: a Review of Quantitative Modeling Approaches." *International Journal of Management Reviews*, 1, 2, 197-224.
Hempel, C.G. 1965. "Science and Human Values." In Hempel, C.G. *Aspects of Scientific Explanation*. New York, NY: The Free Press, 81-96.
Henrich, J., Boyd, R., Bowles, S. Camerer, C. Fehr, E. Gintis, H, McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N.S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., and D. Tracer. 2005. ""Economic man" in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies." *Behavioral and Brain Sciences*, 28, 795-855.
Hinloopen, E., Nijkamp, P., and P. Rietveld. 1983. "The Regime Method: a New Multicriteria Method." In P. Hansen ed. *Essays and Surveys on Multiple Criteria Decision Making*. Berlin: Springer. 146-155.
Hobbes, T. 2002. *Leviathan*, Richard Tuck ed., New York, NY: Cambridge University Press.
Holling, C.S. ed. 1978. *Adaptive Environmental Assessment and Management*. Chichester: Wiley.
Hughes, T.P., Bellwood, D.R., and S.R. Connolly. 2002. "Biodiversity Hotspots, Centres of Endemicity, and the Conservation of Coral Reefs." *Ecology Letters*, 5, 6, 775-784.
Jamieson, D. 2008. *Ethics and the Environment*. New York, NY: Cambridge University Press.
Janzen, D. 1986. "The Future of Tropical Ecology." *Annual Review of Ecology and Systematics*, 17, 305-324.
Jeffrey, R. 1956. "Valuation and Acceptance of Scientific Hypotheses." *Philosophy of Science*, 22, 237-246.
Jost, L., Chao, A., and R.L. Chazdon. 2011. "Compositional Similarity and β (Beta) Diversity." In A.E. Magurran and B.J. McGill, eds. 2011. 66-84.
Justus, J. 2010. "A Case Study in Concept Determination: Ecological Diversity." In K. deLaplante, B. Brown, and K. Peacock, eds. *Handbook of the Philosophy of Science. Volume 11, Philosophy of Ecology*. Waltham, MA: Elsevier. 147-168.
Kagan, S. 1988. "The Additive Fallacy." *Ethics*, 99, 1, 5-31.

Kagan, S. 1997. *Normative Ethics*. Boulder, CO: Westview.
Kant, I. 2002. *Groundwork for the Metaphysics of Morals*. Tr. A. Zweig. New York, NY: Oxford University Press.
Karieva, P. and M. Marvier. 2011. *Conservation Science: Balancing the Needs of People and Nature*. Greenwood Village, CO: Roberts and Company.
Keeney, R.L. 1992. *Value-Focused Thinking*. Cambridge, MA: Harvard University Press.
Keeney, R.L. 2002. "Common Mistakes in Making Value Trade-Offs." *Operations Research*, 50, 6, 935-945.
Keeney, R.L. 2004. "Making Better Decision Makers." *Decision Analysis*, 1, 4, 193-204.
Keeney, R.L. 2007. "Developing Objectives and Attributes." In W. Edwards, R.F. Miles, and D. von Winterfeldt eds. *Advances in Decision Analysis*. New York, NY: Cambridge University Press. 104-128.
Keeney, R.L. and R.S. Gregory. 2005. "Selecting Attributes to Measure the Achievement of Objectives." *Operations Research*, 53, 1-11.
Keeney, R.L. and H. Raiffa. 1993. *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*, 2nd ed. New York, NY: Cambridge University Press.
Kellert, S. 1996. *The Value of Life: Biological Diversity and Human Society*. Washington, DC: Island Press.
Kelly, C. 2008. "The Impossibility of Incommensurable Values." *Philosophical Studies*, 137, 369-382.
Knight, A.T., Cowling, R.M., Difford, M., and B.M. Campbell. 2010. "Mapping Human and Social Dimensions of Conservation Opportunity for the Scheduling of Conservation Action on Private Land." *Conservation Biology*, 24, 5, 1348-1358.
Koricheva, J. and H. Siipi. 2004. "The Phenomenon of Biodiversity." In M. Okansen and J. Pietarinen eds. *Philosophy and Biodiversity*. New York, NY: Cambridge University Press. 27-54.
Ledyard, J. 1995. "Public Goods: A Survey of Experimental Research." In *The Handbook of Experimental Economics*, J. Kagel and A. Roth eds. Princeton, NJ: Princeton University Press. 111-194.
Lemos, N.M. 1994. *Intrinsic Value: Concept and Warrant*. Cambridge: Cambridge University Press.
Lindenmayer, D.B., Thomas, V.C., and T.W. Clark. 1993. "Predictions of the Impacts of Changes in Population Size and Environmental Variability on Leadbeater's Possum, *Gymnobelideus leadbeateri* McCoy (Marsupialia: Petauridae) Using Population Viability Analysis: an Application of the Computer Program VORTEX." *Wildlife Research*, 20, 67-86.
Lindsay, P., Du Toit, J., and M. Mills. 2004. "Area and Prey Reqruiements of African Wild Dogs Under Varying Habitat Conditions: Implications for Reintroductions." *South African Journal of Wildlife Resources*, 34, 77-86.
Lindsay P., Du Toit, J., and M. Mills. 2005. "The Potential Contribution of Ecotourism to African Wild Dog Lycaon Pictus Conservation in South Africa." *Biological Conservation*, 123, 339-348.
Lichtenstein, S. and P. Slovic, eds. 2006. *The Construction of Preference*. New York,

NY: Cambridge University Press.
Luce, M.F. 1998. “Choosing to Avoid: Coping with Negatively Emotion-Laden Consumer Decisions.” *Journal of Consumer Research*, 24, 4, 409-433.
Ludwig, D. 1999. “Is it Meaningful to Estimate a Probability of Extinction?” *Ecology*, 80, 1, 298-310.
Ludwig, D. 2001. “Can We Exploit Sustainably?” In J.D. Reynolds, G.M. Mace, K.H. Redford, and J.G. Robinson, eds. *Conservation of Exploited Species*. New York, NY: Cambridge University Press. 16-38.
Maclaurin, J. and K. Sterelny. 2008. *What is Biodiversity?* Chicago, IL: University of Chicago Press.
Maddock, A. 1999. “Wild Dog Demography in Hluhluwe-Umfolozi Park, South Africa.” *Conservation Biology*, 13, 412-417.
Magurran, A.E. 2004. *Measuring Biological Diversity*. Oxford: Blackwell Science.
Magurran, A.E. and B.J. McGill eds. 2011. *Biological Diversity: Frontiers in Measurement and Assessment*. New York, NY: Oxford.
Margules, C. and S. Sarkar. 2007. *Systematic Conservation Planning*. New York, NY: Cambridge University Press.
McManus, J. 1997. “Tropical Marine Fisheries and the Future of Coral Reefs: a Brief Review with Emphasis on Southeast Asia.” *Coral Reefs*, 16, S121-S127.
McManus, J., Reyes, R.B., and C.L. Nanola. 1997. “Effects of Some Destructive Fishing Methods on Coral Cover and Potential Rates of Recovery.” *Environmental Management*, 21, 69-78.
McShane, T.O., Hirch, P.D., Trung, T.C., Songorwa, A.N., Kinzig, A., Monteferri, B., Mutekanga, D., Van Thang, H., Dammert, J.L., Pulgar-Vidal, M., Welch-Devine, M., Brosius, J.P., Coppolillo, P., and S. O’Connor. 2010. “Hard choices: Making Trade-offs between Biodiversity Conservation and Human Well-being.” *Biological Conservation*, 144, 3, 966-972.
Mellström, C. and M. Johanneson. 2008. “Crowding out and Blood Donation: Was Titmuss Right?” *Journal of the European Economic Association*, 6, 4, 845-863.
Meretsky, V.J., Fischman, R.L., Karr, J.R., Ashe, D.M., Scott, J.M., Noss, R.F., and R.L. Schroeder. 2006. “New Directions in Conservation for the National Wildlife Refuge System.” *BioScience*, 56, 2, 135-143.
Mikalsen, K. and S. Jentoft. 2001. “From User-groups to Stakeholders? The Public Interest in Fisheries Management.” *Marine Policy,* 25, 281-292.
Milgram, E. 2002. “Incommensurability in Perspective.” *Topoi*, 21, 217-226.
Minteer, B. ed. 2009. *Nature in Common?: Environmental Ethics and the Contested Foundations of Environmental Policy*. Philadelphia, PA: Temple University Press.
Mitchell, R., Angle, B., and D. Wood. 1997. “Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts.” *The Academy of Management Review,* 22, 4, 853-886.
Moffett, A. and S. Sarkar. 2006. “Incorporating Multiple Criteria into the Design of

Conservation Area Networks: A Minireview with Recommendations." *Diversity and Distributions,* 12, 125-137.
Moffett, A., Dyer, J.S., and S. Sarkar. 2006. "Integrating Biodiversity Representation with Multiple Criteria in North-Central Namibia using Non-dominated Alternatives and a Modified Analytic Hierarchy Process." *Biological Conservation*, 129, 181-191.
Morgan, G.J. 2010. "Evaluating Maclaurin and Sterelny's Conception of Biodiversity in Cases of Frequent, Promiscuous Lateral Gene Transfer." *Biology and Philosophy,* 25, 603-621.
Myerson, R.B. 2008. "Mechanism Design." In S.N. Durlauf and L.E. Blume eds. *The New Palgrave Dictionary of Economics*. New York, NY: Palgrave Macmillan. 528-542.
Nadasdy, P. 2005. "The Anti-Politics of TEK: The Institutionalization of Co-Management Discourse and Practice." *Anthropologica*, 47, 2, 215-232.
Neige, P. 2003. "Spatial Patterns of Disparity and Diversity of the Recent Cuttlefishes (Cephalopoda) across the Old World." *Journal of Biogeography,* 30, 1125-1137.
Newman, D. and D. Pilson. 1997. "Increased Probability of Extinction Due to Decreased Genetic Effective Population Size: Experimental Populations of *Clarkia Pulchella*." *Evolution*, 51, 2, 354-362.
Norton, B. 1984. "Environmental Ethics and Weak Anthropocentrism." *Environmental Ethics*, 6, 2, 131-148.
Norton, B. ed. 1986. *The Preservation of Species*. Princeton, NJ: Princeton University Press.
Norton, B. 1987. *Why Preserve Natural Variety?* Princeton, NJ: Princeton University Press.
Norton, B. 1994. "On What we Should Save: the Role of Culture in Determining Conservation Targets." In P.L. Forey, C.J. Humphries, and R.I. Vane-Wright, eds. *Systematics and Conservation Evaluation*, New York, NY: Oxford University Press. 23-41.
Norton, B. 2005. *Sustainability: A Philosophy of Adaptive Ecosystem Management*. Chicago, IL: University of Chicago Press.
Norton, B. and D. Noonan 2007. "Ecology and Valuation: Big Changes Needed." *Ecological Economics*, 4, 664-675.
Ogada, M., Woodroffe, R., Oguge, N., and L.G. Frank. 2003. "Limiting Depredation by African Carnivores: the Role of Livestock Husbandry." *Conservation Biology,* 17, 1521-1530.
Okansen, M. 2004. "Biodiversity Considered Philosophically: an Introduction." In M. Okansen and J. Pietarinen eds. *Philosophy and Biodiversity*. New York, NY: Cambridge University Press. 1-26.
Okonta, I. and O. Douglas. 2003. *Where Vultures Feast: Shell, Human Rights, and Oil in the Niger Delta*. New York, NY: Verso.
Organization for Economic Cooperation and Development. 1997. *Towards Sustainable*

*Fisheries: Economic Aspects of the Management of Living Marine Resources*. Paris: Organization for Economic Cooperation and Development.
Osborne, M.J. and A. Rubinstein. 1994. *A Course in Game Theory*. Cambridge, MA: MIT Press.
Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York, NY: Cambridge University Press.
Ostrom, E. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
Ostrom, E., Burger, J., Field, C.B., Norgaard, R.B., and D. Policansky. 1999. "Revisiting the Commons: Local lessons, Global Challenges." *Science,* 284, 278-282.
Pastur, G.M., Lencinas, M.V., Escobar, J., Quiroga, P., Malmierca, L., and M. Lizarralde. 2006. "Understory Succession in Nothofagus Forests in Tierra del Fuego (Argentina) Affected by *Castor Canadensis*." *Applied Vegetation Science*, 9, 1, 143-154.
Payne, J.W. 1976. "Heuristic Search Processes in Decision Making." *Advances in Consumer Research*, 3, 321-327.
Payne, J.W., Bettman, J.R., and D.A. Schkade. 1999. "Measuring Constructed Preferences: Towards a Building Code." *Journal of Risk and Uncertainty*, 19, 1, 243-270.
Pennock, D.S. and W.W. Dimmick. 1997. "Critique of the Evolutionarily Significant Unit as a Definition for "Distinct Population Segments" Under the U.S. Endangered Species Act." *Conservation Biology*, 11, 3, 611-619.
Pinkerton, E., ed. 1989. *Co-operative Management of Local Fisheries: New Directions for Improved Management and Community Development*. Vancouver: University of British Columbia Press.
Rachels, S. 1998. "Counterexamples to the Transitivity of Better Than." *Australasian Journal of Philosophy*, 76, 1, 71-83.
Randall, A. 1986. "Human Preferences, Economics, and the Preservation of Species." In B. Norton, ed. *The Preservation of Species*. Princeton: Princeton University Press. 79-101.
Railton, P. 1986. "Moral Realism." *Philosophical Review*, 95, 163-207.
Rawls, J., 2001. *Justice as Fairness: A Restatement*. Harvard University Press, Cambridge, MA.
Raz, J. 1997. "Incommensurability and Agency," in R. Chang ed. 1997. 110-127.
Raup, D. 1966. "Geometric Analysis of Shell Coiling: General problems." *Journal of Paleontology* 40, 1178-1190.
Redford K.H. and P. Feinsinger. 2001. "The Half-empty Forest: Sustainable Use and the Ecology of Interactions." In J.D. Reynolds, G.M. Mace, Redford, K.H., and J.G. Robinson, eds. *Conservation of Exploited Species*. Cambridge, UK: Cambridge University Press, 370-400.
Reed, M.S. 2008. "Stakeholder Participation for Environmental Management: A Literature Review." *Biological Conservation*, 141, 10, 2417-2431.
Reed, M.S., Graves, A., Dandy, N., Posthumus, H., Hubacek, K., Morris, J., Prell, C.,

Quinn, C.H., and L.C. Stringer. 2009. "Who's in and Why? A Typology of Stakeholder Analysis Methods for Natural Resource Management." *Journal of Environmental Management*, 90, 2009, 1933-1949.

Resnik, M. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis, MN: University of Minnesota Press.

Roberts, C.M., McClean, C.J., Veron, J.E., Hawkins, J.P., Allen, G.R., McAllister, D.E., Mittermeier, C.G., Schueler, F.W., Spalding, M., Wells, F., Vynne, C., and T.B. Werner. 2002. "Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reefs." *Science,* 295, 1280-1284.

Rolston, H., III. 1996. "Feeding People Versus Saving Nature?" In W. Aiden and H. LaFollette, eds. *World Hunger and Morality*. Englewood Cliffs, NJ: Prentice-Hall. 248-267.

Rosenberg, A. 2005. *Philosophy of Social Science*. Boulder, CO: Westview.

Rosenberger, R.S., Peterson, G.L., Clarke, A., and T.C. Brown. 2003. "Measuring Dispositions for Lexicographic Preferences of Environmental Goods: Integrating Economics, Psychology and Ethics." *Ecological Economics*, 44, 63-76.

Rudner, R. 1953. "The Scientist qua Scientist Makes Value Judgments." *Philosophy of Science*, 20, 1-6.

Ryder, O.A. 1986. "Species Conservation and Systematics: the Dilemma of Subspecies." *Trends in Ecology and Evolution*, 1, 9-10.

Sääksvuori, L., Mappes, T. and M. Puurtinen. 2011. "Costly Punishment Prevails in Inter-group Conflict." *Proceedings of the Royal Society B*, 278, 1723, 3428-3436.

Sally, D. 1995. "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958-1992." *Rationality and Society*, 7,1, 58-92.

Salo, A., Keisler, J., and A. Morton eds. 2011. *Portfolio Decision Analysis: Improved Methods for Resource Allocation*. New York, NY: Springer.

Samuelson, P. 1937. "A Note on the Measurement of Utility." *Review of Economic Studies*, 4, 155-161.

Sarkar, S. 2002. "Defining 'Biodiversity'; Assessing Biodiversity." *The Monist,* 85, 1, 131-155.

Sarkar, S. 2005. *Biodiversity and Environmental Philosophy: an Introduction*. New York, NY: Cambridge University Press.

Sarkar, S. 2007. "From Ecological Diversity to Biodiversity." In D. Hull and M. Ruse eds. *The Cambridge Companion to the Philosophy of Biology*, Cambridge: Cambridge University Press, 388-409.

Sarkar, S. 2008. "Norms and the Conservation of Biodiversity." *Resonance*, July 2008, 627-637.

Sarkar, S. 2012a. *Environmental Philosophy: From Theory to Practice*. Malden, MA: Wiley-Blackwell.

Sarkar, S. 2012b. "Complementarity and the Selection of Nature Reserves: Algorithms and the Origins of Conservation Planning, 1980-1995." *Archive for the History of Exact Sciences*. 66, 397-426.

Sarkar, S. and J. Garson. 2004. “Multiple Criterion Synchronization for Conservation Area Network Design: the Use of Non-Dominated Alternative Sets.” *Conservation and Society*, 2, 433-448.

Sarkar, S., Dyer, J.S., Ciarleglio, M., Wang, O., Kemp, N., and C. Margules. 2011. “Developing an Objectives Hierarchy for Multi-Criteria Decisions for Biodiversity Conservation, with a Case Study from Papua, Indonesia.” Unpublished Manuscript.

Scanlon, T.M. 2000. *What We Owe Each Other*. Cambridge, MA: Belknap.

Schiappa, E. 1996. “Towards a Pragmatic Approach to Definition: ‘Wetlands’ and the Politics of Meaning.” In A. Light and E. Katz eds. *Environmental Pragmatism*. New York, NY: Routledge. 209-230.

Schroeder, M. 2012. “Value Theory.” In E. Zalta ed. *The Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*. URL = <http://plato.stanford.edu/archives/sum2012/entries/value-theory/>. Accessed April 2012.

Schüttler, E., Rozzi, R., and K. Jax. 2011. “Towards a Societal Discourse on Invasive Species Management: A Case Study of Public Perceptions of Mink and Beavers in Cape Horn.” *Journal for Nature Conservation*, 19, 3, 175-184.

Seung, T.K. and D. Bonevac. “Plural Values and Indeterminate Rankings.” *Ethics*, 102, 4, 799-813.

Shrader-Frechette, K.S. 2002. *Environmental Justice: Creating Equality, Reclaiming Democracy*. New York, NY: Oxford University Press.

Shrader-Frechette, K.S. and E.D. McCoy. 1993. *Method in Ecology: Strategies for Conservation*. New York, NY: Cambridge University Press.

Singleton, S., 2009. “Native People and Planning for Marine Protected Areas: How ‘Stakeholder’ Processes Fail to Address Conflicts in Complex, Real-world Environments.” *Coastal Management,* 37, 5, 421-440.

Singer, P. 1975. *Animal Liberation*. New York, NY: Random House.

Siurua, H. 2006. “Nature above People: Rolston and ‘Fortress’ Conservation in the South.” *Ethics and the Environment*, 11, 1, 71-96.

Skews, O., Gonzalez, F., Quezada, M., Olave, R., Vargas, V., A. Avila. 1999. “Investigacion Aprovechamiento y Control del Castor en Islas Tierra del Fuego y Navarino.” Technical report, Servicio Agricola y Granadero, XII Region, Punta Arenas.

Slovic, P. 1995. “The Construction of Preference.” *American Psychologist*, 50, 5, 364-371.

Sodhi, N., Acciaioli, G., Erb, M., Tan, K. (Eds.), 2008. *Biodiversity and Human Livelihoods in Protected Areas: Case Studies from the Malay Archipelago*. Cambridge University Press, New York, NY.

Soulé, M. 1985. “What is Conservation Biology?” *BioScience* 35, 727-734.

Spash, C.L. and N. Hanley. 1995. “Preferences, Information, and Biodiversity Preservation.” *Ecological Economics*, 12, 191-208.

Stanley, J. 2005. *Knowledge and Practical Interests*. New York, NY: Oxford University

Press.
Sumaila, U.R. 1999. "A Review of Game-theoretic Models of Fishing." *Marine Policy*, 23, 1-10.
Sunstein, C. 1997. "Incommensurability and Kinds of Valuation: Some Applications in Law." in R. Chang ed. 1997, 234-254.
Swartzendruber, J.F. 1993. *Papua New Guinea Conservation Needs Assessment: Synopsis Report*. Boroko, Papua New Guinea: Papua New Guinea Department of Conservation and Environment.
Takacs, D. 1996. *The Idea of Biodiversity: Philosophies of Paradise*. Baltimore, MD: Johns Hopkins University Press.
Thirgood, S., Redpath, S., Newton, I., and P. Hudson. 2000. "Raptors and Red Grouse: Conservation Conflicts and Management Solutions." *Conservation Biology*, 14, 95–104.
Thompson, D., A. MacDonald, J. Marsden, C. Galbraith. 1995. "Upland Heather Moorland in Great Britain: a Review of International Importance, Vegetation Change and Some Objectives for Nature Conservation." *Biological Conservation,* 71, 163-178.
Titmuss, R. 1970. *The Gift Relationship: From Human Blood to Social Policy*. New York, NY: The New Press.
Trainor, S.F. 2006. "Realms of Value: Conflicting Natural Resource Values and Incommensurability." *Environmental Values*, 15, 3-29.
Tversky, A. and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science*, 211, 453-458.
Udehn, L. 2001. *Methodological Individualism: Background, History and Meaning*. London: Routledge.
U.N. 1948. "The Universal Declaration of Human Rights." URL = <http://www.un.org/en/documents/udhr/>. Accessed November 2011.
U.S. Code. 1973. Title 16. Sections 1531-1544. "Endangered Species Act."
U.S. Department of Health and Human Services. 2012. "Frequently Asked Questions Related to the Poverty Guidelines and Poverty." URL = <http://aspe.hhs.gov/poverty/faq.shtml>. Accessed January 2012.
U.S. Fish and Wildlife Service. 2011. Land Acquisition Priority System. Documents provided by Deborah Holle, Balcones Canyonlands National Wildlife Refuge.
Velland, M., Cornwell, W.K., Magnuson-Ford, K., and A. Mooers. 2011. "Measuring Phylogenetic Biodiversity." In A.E. Magurran and B.J. McGill eds. 2011. 194-207.
Vineberg, S. 2011. "Dutch Book Arguments." In E. Zalta, ed. *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/sum2011/entries/dutch-book/>. Accessed January 2012.
Von Neumann, J. and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
Vucetich, J.A., Nelson, M.P., and M.K. Phillips. 2006. "The Normative Dimension and

Legal Meaning of 'Endangered' and 'Recovery' in the U.S. Endangered Species Act." *Conservation Biology*, 20, 5, 1383-1390.

Waldfogel, J. 1993. "The Deadweight Loss of Christmas." *The American Economic Review*, 83, 5, 1328-1336.

Walzer, M. 1983. *Spheres of Justice*. New York, NY: Basic Books.

Waples, R.S. 1991. "Pacific Salmon, *Oncorhynchus* spp., and the Definition of 'Species' Under the Endangered Species Act." *Marine Fisheries Review*, 53, 11-22.

Weisberg, M. 2007. "Three Kinds of Idealization." *The Journal of Philosophy*, 104, 639-659.

Weitzman, M.L. 1992. "On Diversity." *Quarterly Journal of Economics*. 107, 2, 363-406.

Wenar, L. 2012. "Rights." In E. Zalta, ed. *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/fall2011/entries/rights/>. Accessed March 2012.

White, A.T., Vogt, H.P., Arin, T. 2000. "Philippine Coral Reefs Under Threat: The Economic Losses Caused by Reef Destruction." *Marine Pollution Bulletin,* 40, 598-605.

Wiersma, U. 1992. "The Effects of Extrinsic Rewards in Intrinsic Motivation: A Meta-analysis." *Journal of Occupational and Organizational Psychology,* 65, 101-114.

Wiggins, D. 1997. "Incommensurability: Four Proposals." In R. Chang ed. 1997. 52-66.

Wilson, E.O. ed. 1988. *Biodiversity*. Washington, DC: National Academy Press.

Wood, P.M. *Biodiversity and Democracy: Rethinking Society and Nature*. Vancouver, BC: University of British Columbia Press.

Wright, J., Jones, C., Flecker, A., 2002. "An Ecosystem Engineer, the Beaver, Increases Species Richness at the Landscape Scale." *Oecologia,* 132, 96-101.