

Advancing the diagnostic analysis of environmental problems

Michael Cox

Workshop in Political Theory and Policy Analysis, Indiana University

Abstract: Social-ecological systems exhibit patterns across multiple levels along spatial, temporal, and functional scales. The outcomes that are produced in these systems result from complex, non-additive interactions between different types of social and biophysical components, some of which are common to many systems, and some of which are relatively unique to a particular system. These properties, along with the mostly non-experimental nature of the analysis, make it difficult to construct theories regarding the sustainability of social-ecological systems. This paper builds on previous work that has initiated a diagnostic approach to the analysis of these systems. The process of diagnosis involves asking a series of questions of a system at increasing levels of specificity based on the answers to previous questions. The answer to each question further unpacks the complexity of a system, allowing an analyst to explore patterns of interactions that produce outcomes. An important feature of this approach is the use of multilevel analysis. This paper explores this concept and introduces another – multilevel causation – to further develop the diagnostic approach. It demonstrates that these concepts can be used to analyze a diversity of environmental problems.

Keywords: Causality, diagnostics, panaceas, social-ecological systems

Acknowledgments: I wish to acknowledge support from a subcontract with Dr. Thomas Sterner, University of Gothenburg, Sweden, that is funded by FORMAS.

I. Introduction: complexity and panaceas

1.1. Complexity

Modern society faces a diversity of environmental problems including biodiversity loss, diminishing natural resource stocks, and a changing climate. These problems occur in highly complex systems which are difficult to understand and to manage.

This presents a challenge to scholars and practitioners who desire to understand and effectively address modern environmental problems.

The study of complexity, and the related concept of chaos, has now been popularized across many disciplines (Gleick 2008). Complex systems are composed of diverse parts which strongly interact. Lorenz (1963), for example, modeled a process based on atmospheric convection with a set of three interacting non-linear differential equations. The upshot of the interactions among these equations, along with their non-linearity, was that the system failed to achieve either stability or periodicity, and thus predictability, over any period of time. The system was chaotic. Moreover, minute changes in the starting location of one variable quickly led to drastically different system trajectories. This is now known as sensitive dependence on initial conditions. Similar methods have been applied to ecological systems (Holling 1973; May and McLean 2007). The empirical evidence for chaotic dynamics as it is technically defined is mixed (Hastings 2009). However, it is the case that the behavior of real-world systems is largely non-linear (Stark and Hardy 2003).

There is stronger empirical support, particularly in aquatic systems, for similarly self-magnifying processes in the form of cascading ecological trophic effects (Borer and Gruner 2009). These can occur when the removal or suppression of a keystone species affects other species and environmental parameters (Scheffer et al. 2005). Similarly, Ostrom (2005) notes that sets of rules in governance contexts interact in a complex manner to configuratively produce an outcome. Removing one rule can change the effects of other rules.

As a result of the intricate interactions between a set of variables, their contributions to the progression of the system and to outcomes are non-additive. In producing an outcome they form a web of causal relationships, where pulling on one thread can affect the rest of the web. When a change in one variable can affect the larger system through its interactions with other variables, it becomes extremely difficult to predict what will occur after such a change. Similarly, it is difficult to generalize the effects of any one variable from one system to other systems, because its effects could be contingent upon factors specific to that system. Some degree of generalizability and predictability is required, however, if we are to make useful prescriptions on how to interact with such complex systems.

1.2. Panaceas and simplification

In this context, we can expect the governance of complex systems to be highly challenging. Unfortunately, the history of natural resource governance has been typified by a focus on simplified solutions, or panaceas (Meinzen-Dick 2007; Ostrom 2007, 2009; Ostrom and Cox 2010). The two most common embodiments of such panaceas are private and public property regimes, such as national parks in the case of public ownership and water markets when they create private rights. Simplified solutions can lead to the failure of governance arrangements to address

the diversity of contexts to which these solutions are applied (Korten 1980; Scott 1998).

In addition to, and partly as a result of, this simplified management landscape, a common result in natural resource management has been a reduction in the complexity of natural systems. Simplification is a ubiquitous feature of human-environment interactions, made most visually obvious by monocultures in agricultural systems. Other examples include highly engineered approaches to water management, most commonly in the form of dams, and fire suppression in forest ecosystems. Holling and Meffe (1996) argue that such simplification, and the suppression of variability through control, harms the long-term viability and robustness of such systems.

More recently, there has been a movement to recognize the diversity of governance arrangements, echoing the modern emphasis on biological diversity (Ostrom 2005). In the literature focusing on medium-to small-scale common-pool resource management, this movement has led to the identification of a large set of variables that have been shown to affect natural resource outcomes (Agrawal 2003). The identification of so many variables, however, returns us to the problem of achieving predictability and generalizability in the face of their complex interactions. We cannot simply analyze a subset of the potentially relevant variables without worrying that our results will be biased by the interactions between these variables and others that we have not included in our analysis. Meanwhile, measuring all potentially relevant variables is infeasible in any given study.

The challenge we currently face is to construct tractable ways to analyze complex systems without appealing to overly-simplified solutions. One response to this challenge is a diagnostic approach to social-ecological analysis, and with respect to this approach, this paper has two goals. The first is to further develop the conceptual basis of a diagnostic approach, building on previous work in the philosophy of science. The second and primary goal of this paper is to show how the diagnostic approach can help an analyst deal with complexity when examining a particular case or problem.

2. The diagnostic approach

The diagnostic approach has been described by Young (2002, 176):

The diagnostic approach seeks to disaggregate environmental issues, identifying elements of individual problems that are significant from a problem-solving perspective and reaching conclusions about design features needed to address each of the elements identified.

The motivation for diagnostic thinking is to devise responses to problems to which they are appropriately tailored. Implicit here is the empirically well-established premise that no one solution can solve all problems. To avoid the panacea problem of overly simplistic prescription and confront complexity, we want a body of

knowledge that can enable us to address a diversity of individual cases in a way that recognizes the prescriptive implications of both their similarities and their differences.

Informally, the concept of diagnosis has meant to ascertain the cause of a problem, with the idea that by intervening with the cause, the problem could be ameliorated. In this paper to diagnose an outcome also means to ask a series of questions of a particular case, with: (1) subsequent questions being asked based on the answers to previous questions, and (2) subsequent questions regarding variables that are more specific to a subset of cases. These questions are asked in order to achieve the following three goals: (1) to identify the causes of a particular outcome in a case; (2) to compare this case to others as a means of deriving generalizations or theories about a set of cases; (3) to use this knowledge to formulate hypothesis or prescriptive predictions. These three goals roughly correspond to three methods of scientific reasoning: abduction, induction, and deduction (see Bromley 2006). Each deals with three types of statements: a rule, a fact about a case, and a result.

The first step of the diagnostic process is to observe an outcome and explore the conditions that could lead to it. This is sometimes referred to as abductive reasoning. It begins with a rule statement and a (potentially surprising) result for a case, and stipulates a fact of the case that relies on the rule to explain the result. Abduction as generally described takes a particular form, as demonstrated by this example:

If it rains, the lawn gets wet; (Rule)
The lawn is wet; (Result)
Therefore, it rained last night (Case)

The rule statement can be similar to a theory, and it has two components: an antecedent and a consequent, with the former predicting the latter. The result states an observation to be explained. The case statement explains the result. The first two statements are the premises, and the third is the conclusion that follows. If, as has been traditionally understood, the rule is stating that the antecedent is sufficient, but not necessary, for the consequent (rain is sufficient for a wet lawn), then abduction only reveals a candidate explanation. To definitively conclude the presence of a fact in a case from a rule and result is to commit a logical fallacy, because other facts could explain the result (someone could have watered the lawn). Abduction is distinguished from deduction by the fact that deductive reasoning does not lead to such a logical fallacy because its conclusions necessarily follow from its premises. As generally understood, the following (commonly referred to as *modus tollens*) is a valid deductive argument:

If it rains, the lawn gets wet; (Rule)
The lawn is not wet; (Result)
Therefore, it did not rain last night (Case)

If the rule in this argument is stating that rain is sufficient for a wet lawn, then this is a deductive (necessary) conclusion. Due to the complexity previously described,

however, no single variable is likely to be sufficient for an outcome. Many other factors also matter. Thus, although these two arguments would traditionally be distinguished as being abductive and deductive respectively, in this discussion I will consider them to be of one type, because they both seek to explain a result by stipulating a fact about a case. I will refer to this as result-based reasoning, because it begins by considering a result that it then seeks to explain.

To help us produce the rule statements that are assumed in result-based reasoning, we can rely on inductive reasoning. It has the following general form:

It rained last night; (Result)
 The lawn is wet; (Case)
 Therefore, if it rains, the lawn gets wet (Rule)

Following the same naming scheme, we can call this case-based reasoning (reasoning from cases). Such reasoning infers a general pattern or theory from a set of observations.

Once we have hypothesized that a condition is important in producing an outcome, we can use result-based reasoning, and we can also use what we can call rule-based reasoning. This is similar to what is commonly called deduction. The most common example is this type of reasoning is the following argument, known as *modus ponens*.

If it rains, the lawn gets wet; (Rule)
 It rained last night; (Case)
 Therefore, the lawn is wet (Result)

If the combined set of premises is true, and the rule is describing a relationship in which the antecedent is sufficient for the consequent, then the truth of the result can be automatically inferred. However, as stated earlier, in studying complex systems it is unlikely that we will have many cases where we know that a single condition would be sufficient to obtain an outcome. Therefore in lieu of the term deductive, which implies a necessary relationship between premises and a conclusion, I will follow the same scheme as before and refer to it as rule-based reasoning.

A primary use of rule-based reasoning is to provide prescriptions to cases similar to those that were used to produce the theory given in the rule statement. Rule-based reasoning predicts a result from the combination of one or more facts of a case and one or more rule statements. This gives us the ability to predict, and therefore, prescribe, which is the final goal of diagnostic analysis.

The standard approach in forming institutional prescriptions in mainstream economics is similar to this, in the sense that it uses rule-based reasoning. Specifically, the traditional approach uses deductive syllogisms to derive prescriptive conclusions automatically from a set of rules and facts about a case. In traditional economics the rule statements of a deductive syllogism are derived *a priori*.

A basic example of rule-based reasoning in mainstream economics is the prescription of certain arrangements for managing certain types of goods. Private

goods are automatically seen as being better managed and distributed by market mechanisms, whereas public goods are seen as being better managed and governed by public organizations or governments. The deductive process for a good like water might go as follows:

Private goods that are managed by well-functioning markets or allocated efficiently;

Water is a private good;

Therefore, water will be allocated efficiently by a well-functioning market

The source of rule statements is where the diagnostic approach, at least as I am describing it, diverges from the traditional approach. The goal of the diagnostic approach, as stated by Young, is to match institutional prescriptions to social and biophysical diagnostic criteria. The traditional approach does have a process by which it matches criteria to a particular prescription. However, the range of prescriptions has been rather simplified in comparison with the diversity of contexts to which these prescriptions have been applied. This is not necessarily surprising, given that the policy sciences have yet to undergo a process of observation and taxonomic classification of comparable breadth and depth to the one that began long ago in the biological sciences, most famously with the work of Carl Linnaeus. Scholars' knowledge of the diversity of life, while by no means complete, is at this point quite substantial and taxonomically rigorous. There has not been a process of comparable structural rigor in the study of environmental policy and social-ecological systems that could be used to build knowledge relating institutions and other factors from the bottom up.

It is worth noting here that the only form of reasoning just described that does not take the rule statement as given is induction, or case-based reasoning. This implies that in order to both generate and to test theories and predictions, we need to use case-based reasoning, which is consonant with the emphasis on observation in the work on biological classification. In the diagnostic approach then, the rule stating that private goods are most efficiently managed through markets would be established by observing many cases where this occurs (through case-based reasoning). In a manner similar to diagnostic medical expertise (Ericsson 2003), the generation of diagnostic knowledge and expertise in environmental management then is largely based on experiences with numerous cases.

To conclude, we can return to the concept of diagnosis with these three basic types of reasoning in mind. Diagnosis *of a specific case* begins with result-based reasoning, using existing theories and an outcome or result to be explained. These theories are established using case-based reasoning, comparing many cases with each other. For example, we might have a theory stating that monitoring the condition of a resource contributes to sustainable forest management. If we find a case where tree harvesting is unsustainable, then we might look to see whether there is no forest monitoring (this follows the logic of *modus tollens* described earlier). The final step in diagnosis is to use rule-based reasoning to prescribe a change, if the observed outcome is undesirable.

The scientific method and the specific process of diagnosis as carried out through these three types of reasoning requires that we ask a variety of questions. The challenge is to structure these questions in a way that does not overwhelm the resources of the analyst. There are simply too many questions that are potentially relevant to any given case for all of them to be asked. And yet if we do not include one, this could bias our results. One reason that the traditional rule-based economic approach has been so popular is that it limits the number of questions that are asked. But, in so doing, it has often led to overly simplified governance prescriptions.

3. Dealing with complexity by diagnosing social-ecological systems

3.1. Multiple levels

Diagnosis involves a sequence of questions, whereby subsequent questions build on more fundamental questions. We face the challenge of structuring these questions in a way that guides an analyst from very general questions, which are relevant to most systems, to more specific questions, that are relevant to and asked of smaller subsets of systems. In presenting a basic framework for the analysis of social-ecological systems (SESs), Ostrom (2007, 2009) provides one response to this challenge. This framework can guide an exploration of a case by organizing the relevant features of that case along multiple levels. The framework contains several levels of analysis. The first level aggregates the relevant features of a SES into four main components: resource systems, resource units, governance systems, and actors (see Figure 1). The second level contains features of a SES that Ostrom associates with each these components, and subsequent levels further unpack these properties into features that are more specific to subsets of cases (see Brock and Carpenter 2007; Meinzen-Dick 2007; Ostrom and Cox 2010) (see Figure 2). Recent work has been done to further develop and implement the SES framework (Basurto and Ostrom 2009; Blanco 2010; Fleischman et al. 2010; McGinnis 2010).

For this discussion, the key feature of this work is the presence of multilevel analysis, with levels varying in their degree of specificity to individual cases. Subsequent levels ideally describe variables that are more specific to a smaller subset of cases. A multilevel arrangement like this is important for a diagnostic approach, which asks a series of questions of a system, with subsequent questions becoming increasingly specific based on answers to previous questions. Multiple levels of analysis could structure such a series of questions.

The answer to each diagnostic question assigns a case to a subpopulation of cases, about which there are some subsequent questions that are likely to be relevant, and others that are not. For example, resource system “sector” is a high level variable in the SES framework, distinguishing between systems based on the type of resource being managed. There are questions we would ask of groundwater

Revised SES Framework With Multiple First-Tier Components

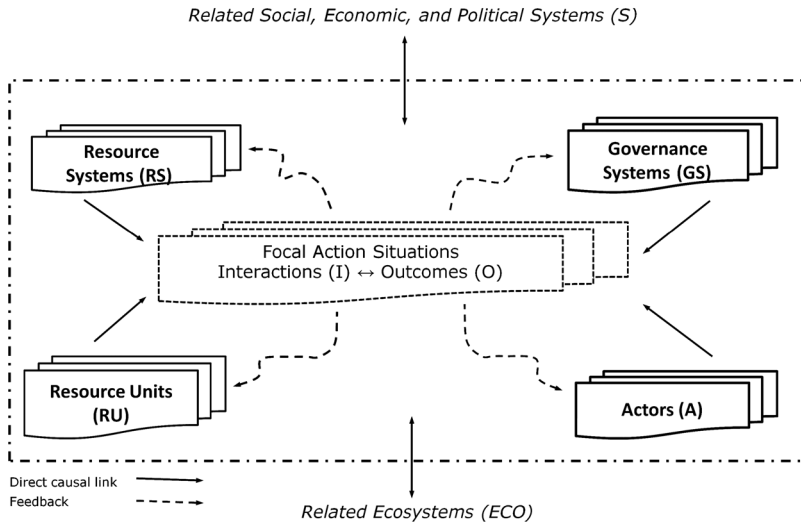


Figure 1: SES framework first tier components (Source: Ostrom and Cox 2010).

system, such as the level of salinity and soil properties, that would be less salient when examining marine fishery systems.

The increasing specificity of subsequent questions based on answers to previous questions, and the concomitant typological decomposition of the system under analysis, is important for several reasons. First, it helps avoid the panacea problem of overly simplified institutional prescriptions by facilitating the construction of theories at multiple levels of specificity/generality, and specific to

| Resource Systems | Resource Units | Governance Systems | Actors | Action Situations |
|---|---|--|---|---|
| <ol style="list-style-type: none"> 1) Sector 2) Boundary clarity 3) Size <ol style="list-style-type: none"> a) Area b) Volume 4) Infrastructure 5) Productivity 6) Equilibrium properties <ol style="list-style-type: none"> a) Recharge dynamics b) Recharge rate c) Number of equilibria d) Feedbacks <ol style="list-style-type: none"> i) Positive ii) Negative 7) Predictability 8) Storage capacity 9) Location | <ol style="list-style-type: none"> 1) Resource unit mobility 2) Replacement rate 3) Interactions <ol style="list-style-type: none"> a) Strong to weak b) Predatory or symbiotic 4) Economic value 5) Size <ol style="list-style-type: none"> a) Large to small b) Trophic level 6) Distinctive markings 7) Distribution <ol style="list-style-type: none"> a) Spatial heterogeneity b) Temporal heterogeneity | <ol style="list-style-type: none"> 1) Rules <ol style="list-style-type: none"> a) Operational rules b) Collective-choice rules c) Constitutional rules 2) Property-rights regime <ol style="list-style-type: none"> a) Private b) Public c) Common d) Mixed 3) Network structure <ol style="list-style-type: none"> a) Centrality b) Modularity c) Connectivity d) Number of levels | <ol style="list-style-type: none"> 1) Group size 2) Socioeconomic attributes <ol style="list-style-type: none"> a) Economic b) Cultural 3) History of use 4) Location 5) Leadership 6) Social capital 7) Knowledge of SES 8) Resource dependence 9) Technology used | <ol style="list-style-type: none"> 1) Process <ol style="list-style-type: none"> a) Monitoring <ol style="list-style-type: none"> i) Environmental ii) Social b) Sanctioning c) Conflict resolution d) Provision <ol style="list-style-type: none"> i) Informational ii) Infrastructural e) Appropriation f) Policymaking |

Figure 2: SES framework second tier variables (Source: Ostrom and Cox 2010).

subsets of cases, as is warranted by the data and information available. Second, it helps to deal with the potentially overwhelming complexity of SESs. Addressing the panacea problem by using multiple levels of analysis and aggregation is the topic of several other papers (Cox 2008; Ostrom and Cox 2010). Here I will focus on how the diagnostic approach can confront complexity, leaving the full integration of these two topics to future work.

3.2. Limited diversity

The diagnostic approach as currently outlined would be difficult to apply if it weren't for the fact that SESs display what Ragin (2000) refers to as "limited diversity." This means that the variables describing the state of a SES are not independent, but mutually constrain each other. Ragin (2000) is referring to social systems, but the same can be said for SESs. For example, two variables describing the populations of a predator and its prey will mutually constrain each other through bottom-up and top-down regulation (Scheffer et al. 2005). A common observation made of complex systems is that they are characterized by particular regimes or basins of attraction, in which many of the variables constrain the values that others can take on (Holling 1973; Walker et al. 2006).

Whenever there is a systematic relationship between variables of a system, this will tend to constrain them to a subset of their possible values, and the system will have fewer degrees of freedom than it has variables describing its state. Thus, while SESs have numerous variables that produce explosively large numbers of possible configurations of values, in reality we may only have to explore a small subset of these dimensions. Moreover, the inductively discovered patterns between variables in one set of systems can be used to construct hypotheses regarding the likely values of some variables based on the values of others in another set of systems.

For the diagnostic process, we also need to resort to another type of limited diversity. The set of variables that is relevant for any particular SES is very limited in comparison to the set of variables that are potentially relevant for all SESs. For example, if an analyst is examining the use of a forest ecosystem, he or she will likely need to ask some questions about ecological interactions and other important biophysical features such as soil properties (Tucker et al. 2007). However, if he or she is examining a groundwater aquifer system or an atmospheric pollution problem, ecological interactions may become less important, while implementing environmental monitoring of the system may become more technologically complicated, giving rise to an alternative set of questions and relevant variables. All potentially relevant variables need not be explored for each case.

To take advantage of this type of limited diversity, we need a way to structure a set of variables such that the values of those that are more general give some indication as to whether or not others are likely to be relevant to our case. One way to do this is to arrange variables across levels such that variables at each level aggregate variables at the next level. This takes advantage of a principle stated

by Levin (1999, 63), that “the coarser the field of vision, the easier it is to make definitive statements” over a range of cases. As we aggregate, the statements made with the more aggregated concepts will tend to be more accurate across a range of cases. Thus, if we want to arrange variables in a way that prioritizes those that are most likely to be useful in making accurate statements for a broad diversity of systems, this is a useful design. Beginning with these, we can proceed to their more disaggregated counterparts.

An aggregated variable could be a general type of more specific subtypes at a more disaggregated level. To use the previous example, the type of resource system is an aggregated variable. Based on the value this variable takes on, we know to ask different subsequent questions about particular variables because different types of resource systems have some properties associated with them which others do not (for example, groundwater and the atmosphere resource systems have a volume while a forest does not). At each successive level of disaggregation, the statements we make will be more specific to a smaller subset of systems, in accordance with Levin’s principle.

4. Multiple levels of causation

There is an ongoing effort to develop multilevel analysis using the type-subtype relationship in order to address the panacea problem and confront complexity. In addition to using the type-subtype relationship, we can aggregate a set of variables by a common distal or ultimate cause they share. Here I will focus on this type of aggregation.

Levels of causation are generally arranged from being in some way closest to the outcome of interest (proximate causes), to being more distanced but frequently more encompassing or complete in their explanatory power (distal causes). A basic question that a diagnostician might ask is, what are the proximate vs. distal causes of a particular outcome or problem? To fully understand an outcome, we need to understand both types of explanations, and the complex webs of causation that they evoke. Additionally, in different situations ameliorating a problem may be more easily or effectively done by addressing either a relatively proximate or more distal cause.

Several disciplines have introduced multiple levels of causation to sort out explanations of complex phenomena. Within the discipline of biology and research in evolutionary processes, a distinction between proximate and ultimate causes has been most popularized by Mayr (1961, 1988). In this context, a proximate cause explains the mechanism an organism has adopted to achieve a particular outcome, and an ultimate cause explains the proximate cause in evolutionary terms, or why this behavior or quality is evolutionarily adaptive. For example, Stephenson (1981) first explains how a variety of plants are able to shed a portion of their developing flowers and fruits (proximate cause), and then discusses why this may be an evolutionarily adaptive behavior (ultimate cause).

Ariew (2003, 564), using evolutionary instead of ultimate as his preferred term, gives a description of the distinction that is useful for this discussion:

Reference to proximate causes answer various questions including, ‘How does something get built?’ and ‘How does something operate?’. Evolutionary explanations (which substitute Mayr’s ‘ultimate cause’) ... are statistical explanations that refer to ensemble-level events that track trends in populations rather than the vagaries of individual-level causal events. By averaging out individual-level differences, evolutionary explanations pick out patterns in common to all evolutionary events.

A similar distinction has arisen within the public health and epidemiological literature, examining proximate vs. distal causes of risk factors for disease. Here, proximate causes refer to those individual-level features that explain a health outcome for an individual, while distal causes refer to population-level features that are statistically related to the rate of incidence of diseases for a population. For example, a proximate cause for a heart attack could be hypertension, while a more distal cause could be socio-economic status (Link and Phelan 1995).

Scholars studying processes involved in land change, particularly deforestation, have also found the proximate-distal distinction useful (Ojima et al. 2001; Geist and Lambin 2002; Carr 2004; Turner and Robbins 2008). Based on a meta-analysis of 156 cases, for example, Geist and Lambin (2002) list several proximate causes of tropical deforestation (agricultural expansion, wood extraction, and infrastructure) and several underlying causes (including market growth, government policies, and technological change).

Finally, in discussing the possible causes of deforestation, Bromley (2006) makes a distinction between mechanical and teleological causes. Here, a mechanistic, or proximate, cause of deforestation might be the construction of a new road near a forest. A teleological cause of deforestation would be the interests that are “served by deforestation and the conversion in land cover and land use, and how those interests manage to work the political system so that their purposes are achieved” (Bromley 2006, 171). In this context a teleological cause is similar to an evolutionarily ultimate cause, in that sense that it describes the purpose that is served by a more proximate cause. The purpose here, however, is not presumed to be evolutionarily adaptive. Instead, Bromley has in mind the human motivations and incentives behind the more proximate causes. Bromley likens the proximate causes to symptoms of the more fundamental teleological causes.

5. How multicausal analysis can aid in diagnosis

Here I discuss how proximate and distal causes have implicitly been used in previous research. I briefly explore two research programs – one on community-based management of common-pool resources, and the other, policy instrument choice for environmental pollution problems. The common thread in these two

examples is the aggregation from proximate causes to distal causes, which could be then used as the starting point for a diagnostic analysis of a case or problem.

5.1. Community-based management

The study of community-based natural resource management systems is a well developed research program. One strand has occurred at the Workshop in Political Theory and Policy Analysis at Indiana University. This has involved several projects that developed and analyzed databases of a large number of community-based systems (Ostrom 1990; Schlager 1990; Tang 1992; Lam 1998; Agrawal 2005; Hayes 2006; Ostrom and Nagendra 2007; Cox et al. 2010). The primary goal of these endeavors was to explain the persistence of some community-based natural resource management regimes.

One result of the first of these endeavors was Ostrom's principles for sustainable community-based natural resource management (1990). Ostrom had searched for specific rules that characterized successful long-term community-managed systems. However, not finding any that were consistent across the observed cases, Ostrom (1990) turned to a higher level of aggregation, which she referred to as design principles. The principles stipulated broad conditions, such as the presence of accountable monitoring, conflict resolution mechanisms, and graduated sanctions.

As noted by Wilson (2010), Ostrom's design principles can be seen as distal causes:

Ostrom initially attempted to correlate the success of each group with specific proximate mechanisms. Because there are many ways to skin a cat, the proximate mechanisms that work successfully in one group need not operate in other successful groups, resulting in weak correlations in a statistical analysis. When Ostrom started to focus on design principles, she was studying ultimate causation. The design principles are required for success, no matter how they are implemented, resulting in strong correlations. Of course, studying or advising any particular group would require close attention to both ultimate and proximate causation.

In developing the design principles, Ostrom (1990) was relying on case-based reasoning. The value of this process is that we can use rule-based reasoning to formulate prescriptions to cases similar to those from which the principles were developed. Indeed, Ostrom's design principles have been criticized as a new panacea of the type described earlier (Cox et al. 2010). Using the proximate-distal distinction helps us to examine this issue. As Wilson indicates, "there are many ways to skin a cat." Each principle can be satisfied by a range of proximate conditions, and thus a prescription for an institutional design principle is not as constraining as a more proximate prescription (a particular rule) would be.

Ariew (2003) offers a distinction between individual fitness and trait-fitness that is helpful here, where individual fitness results from proximate causes, and

trait-fitness relates to distal causes. The design principles confer trait-fitness, or the average robustness of the systems that have these traits. This statistical, or non-deterministic, quality is important to emphasize. In their review of the literature written since the introduction of the principles, Cox et al. (2010) find that none of the principles guarantee success in their presence or failure in their absence. Wilson's language above is thus too strong.

Understanding the design principles as distal causes has an important diagnostic implication. Generally, we can use relatively distal variables such as those embodied in the principles as a way approaching a novel case diagnostically, by beginning with questions about each of the principles. Then, depending on answers to these questions, we could conceivably proceed to ask questions about the more specific (proximate) conditions that could be facilitating these conditions. The point is that we do not need to approach each new case by asking about innumerable sets of specific rules relevant to natural resource management, even if there are many such rules that are potentially relevant. We can use a diagnostic process to guide us to ask questions about particular rules that appear likely to be relevant under particular circumstances.

5.2. Policy instruments and climate change

In contrast with the commons literature, which has tended to emphasize small-scale community-based systems, the literature on environmental economics and policy instrument choice has focused mostly on larger-scale environmental pollution problems. As alluded to earlier, this program has taken somewhat of a diagnostic approach by asking a series of questions, the answers to which have implications for the choice of the best policy instrument. The initial set of diagnostic questions asked are as follows: first, is there evidence of a market failure that leads to a divergence between private and social interests? Secondly, what is the type of failure? Is it, for example, a positive or negative externality, imperfect competition, information asymmetry, or a monopoly? The institutional design implications for each type of failure differ, as they should in a diagnostic approach.

One of the major issues that environmental economists focus on is global warming as a result of the emissions of greenhouse gases (GHGs), particularly CO₂. We can better understand the issue of CO₂ emissions and other GHGs and their implications for climate change with multilevel causation. Within the economic literature on climate change, there is a distinction between adaptation and mitigation (Stern 2007). Mitigation generally involves attempts to lower the amounts of GHGs contained in the atmosphere or to otherwise shift its energy balance, while adaptation involves adapting to the effects of higher temperatures and other changes that will likely result from a strengthened greenhouse effect.

Mitigation efforts include policy-based methods such as cap-and-trade systems, renewable energy subsidies or reforestation policies, as well more technology-based geo-engineering methods such as aerosol emission, oceanic cloud-seeding,

and carbon sequestration. Adaptation methods are specific to the effects of climate change on systems that have been identified as socially or ecologically important. These impacts include changes in long-term averages of temperature or water availability and increasingly large fluctuations about those averages, changing species distributions, and rising sea levels.

These two approaches, adaptation and mitigation, address the proximate and the distal causes of a variety of problems. This application is most similar to the way these concepts are used in the land-use change literature mentioned earlier. For example, invasive species are a problem in many parts of the world. These can be considered a proximate cause for the disruption of local ecologies and food webs. In some of the affected systems, a more distal cause is a changing climate that results from an increased greenhouse effect. With warmer local and regional temperatures come modified ranges of species which may enter previously inhospitable regions. For some systems then, invasive species are one symptom of a more general problem, just as proximate causes of poor health can be considered to be symptoms of broader socio-economic problems.

We could also consider what are the distal causes of greenhouse gas (GHG) emissions. This would involve what Bromley (2006) referred to as teleological causes. This would lead to a whole new set of questions, depending on the pollutant that were considering. If the pollutant is CO₂, then we have to ask what interests are served by the combustion of fossil fuels to more thoroughly unpack the reasons ultimately behind the introduction of new species into an ecosystem. If the GHG is methane, we also must ask about combustion, but also about other important factors such as industrial agriculture and livestock operations and landfills. Identifying the interests that are served by these activities gives us at least an analytical point of leverage for addressing the initial problem were concerned with – invasive species.

5.3. Conclusions from the two examples

There are several points we can make with these two examples now in mind. If we want to start a diagnostic analysis with a cause that is general to many systems, we could usefully start with a relatively distal cause. At the same time, these cases illustrate that reasoning from proximate to distal causes is also important. This is not diagnostic thinking as I have defined it, which goes from more general conditions to more specific conditions. However, these examples show that reasoning from specific to general conditions is a critically important process to enable the diagnostic analysis of any particular case. This is largely case-based (inductive) reasoning used to produce rule-statements.

The climate change example demonstrates another important point: just because a more distal cause encompasses many more systems, this does not mean that this is the best leverage point for improving conditions in local system. In this case, changing a distal cause, such as the warming climate, is much more difficult to do for a group of actors than trying to help a particular ecosystem adapt

to a new invasive species. This is analogous to the difference between addressing a distal cause of poor health, such a socio-economic status, and intervening to affect a variety of proximate causes of individual outcomes. While doctors may be aware that intervening in the distal causes of many of the maladies they treat would ultimately be more effective for the larger population, they do not have the resources to change the socio-economic status of their patients or the larger socio-economic fabric that may contribute to the health problems of their patients.

These two examples are very different, in large part because of the contrasting scales of the systems that are addressed. However, what we see from each example is that employing multiple levels of causality can be a useful way to structure a sequence of questions that could in turn be used to approach a system or problem diagnostically. Such an approach enables an analyst to choose the most likely relevant questions to ask, among the potentially innumerable number of questions that could be asked at any point in an analysis. It then can guide them to ask subsequent questions, based on the answers to the initial questions.

6. Conclusions and challenges

Adopting a more problem-based, diagnostic approach to research in environmental policy and management is complicated by several factors. First, the systems are not amenable to experimental analysis. Second, ecosystems and SESs have a large number of state variables. Third, the costs of collecting data on these variables for a SES can be very high. Fourth, SESs frequently exhibit low levels of feedback and substantial hysteresis, which make learning and adaptation very difficult. Global-scale problems with high levels of irreversibility such as climate change do not offer much opportunity for diagnostic learning (case-based reasoning) at the global scale. For this reason, and for others, producing and maintaining a high level of institutional diversity will greatly facilitate the diagnostic approach. Global scale problems can be addressed as multiple scales, and smaller scales offer more opportunities for experimentation, learning, and pattern recognition.

Another challenge to the diagnostic approach is the fact that the values of important variables, which one uses as a basis for an institutional or technological prediction and prescription, may be endogenous to changes made in implementing this very prescription. This can result from the strategic behavior of human actors in SESs, as well as complex ecological interactions between ecological agents. Holling and Meffe (1996, 330) refer to a failure to recognize this endogeneity as the pathology of natural resource management, and argue that negative outcomes can be expected from “a command-and-control approach to renewable resource management, where it is believed that humans can select one component of a self-sustaining natural system and change it to a fundamentally different configuration in which the adjusted system remains in that new configuration indefinitely without other, related changes in the larger system.” They cite pest outbreaks and excessive

forest fires as examples of such negative outcomes. Endogeneity of supposedly independent variables is a problem for causal inference in social science generally (King et al. 1994). In recognizing this endogeneity, the diagnostic approach needs to be consonant with the research program in adaptive environmental management (Holling 1978).

To conclude, this paper is highly exploratory, and much additional work will need to be done to clarify the concepts and their relationships, and to test these with empirical work. In spite of these challenges and the work that remains, hopefully this paper has illustrated that a diagnostic, problem-based approach to research in this field has much to offer as an organizing logic for addressing a diversity of environmental problems.

Literature cited

- Agrawal, A. 2003. Sustainable governance of common-pool resources: Context, methods, and politics. *Annual Review of Anthropology* 32:243–262.
- Agrawal, A. 2005. *Environmentality: Technologies of Government and the Making of Subjects*. Durham, NC: Duke University Press.
- Ariew, A. 2003. Ernst Mayr's 'ultimate/proximate' distinction reconsidered and reconstructed. *Biology and Philosophy* 18:553–565.
- Basurto, X. and E. Ostrom. 2009. Beyond the tragedy of the commons. *Economia delle fonti di energia e dell'ambiente* 52(1):35–60.
- Blanco, E. 2010. A social-ecological approach to voluntary environmental initiatives: The case of nature-based tourism. *Policy Sciences* 44(1):35–52.
- Borer, E. T. and D. S. Gruner. 2009. Top-down and bottom-up regulation of communities. In *The Princeton Guide to Ecology*, ed. S. A. Levin. Princeton, NJ: Princeton University Press.
- Brock, W. A. and S. R. Carpenter. 2007. Panaceas and diversification of environmental policy. *PNAS* 104(39):15206–15211.
- Bromley, D. 2006. *Sufficient Reason: Volitional Pragmatism and the Meaning of Economic Institutions*. Princeton, NJ: Princeton University Press.
- Carr, D. L. 2004. Proximate population factors and deforestation in tropical agricultural frontiers. *Population and Environment* 25(6):585–612.
- Cox, M. 2008. Balancing accuracy and meaning in common-pool resource theory. *Ecology and Society* 13(2).
- Cox, M., G. Arnold, and S. Villamayor-Tomas. 2010. A review of design principles for community-based natural resource management. *Ecology and Society* 15(4).
- Ericsson, K. A. 2003. Deliberate practice and acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine* 79(10):S70–S81.
- Fleischman, F. D., K. Boenning, G. A. Garcia-Lopez, S. Mincey, M. Schmitt-Harsh, K. Daedlow, M. Claudia Lopez, X. Basurto, and B. Fischer. 2010. Disturbance, Response, and Persistence in Self-Organized Forested Communities: Analysis

- of Robustness and Resilience in Five Communities in Southern Indiana. *Ecology and Society* 15(4).
- Geist, H. J. and E. F. Lambin. 2002. Proximate causes and underlying driving forces of tropical deforestation. *Bioscience* 52(2):143–150.
- Gleick, J. 2008. *Chaos: Making a New Science*. Penguin.
- Hastings, A. 2009. Biological chaos and complex dynamics. In *The Princeton Guide to Ecology*, ed. S. A. Levin. Princeton, NJ: Princeton University Press.
- Hayes, T. 2006. Parks, people, and forest protection: An institutional assessment of the effectiveness of protected areas. *World Development* 34(12):2064–2075.
- Holling, C. S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics* 4:1–23.
- Holling, C. S. 1978. *Adaptive Environmental Assessment and Management*. London, UK: John Wiley & Sons.
- Holling, C. S. and G. K. Meffe. 1996. Command and control and the pathology of natural resource management. *Conservation Biology* 10(2):328–337.
- King, G., R. O. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Korten, D. C. 1980. Community organization and rural development: A learning process approach. *Public Administration Review* 40(5):480–511.
- Lam, W. -F. 1998. *Governing Irrigation Systems in Nepal: Institutions, Infrastructure and Collective Action*. Oakland, CA: ICS Press.
- Levin, S. 1999. *Fragile Dominion: Complexity and the Commons*. Cambridge, MA: Perseus Books.
- Link, B. G. and J. Phelan. 1995. Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior* 35:80–94.
- Lorenz, E. N. 1963. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20:130–141.
- May, R. and A. McLean. 2007. *Theoretical Ecology: Principles and Applications*. Oxford, UK: Oxford University Press.
- Mayr, E. 1961. Cause and effect in biology. *Science* 143:1501–1506.
- Mayr, E. 1988. *Toward a New Philosophy of Biology: Observation of an Evolutionist*. Cambridge, MA: Harvard University Press.
- McGinnis, M. 2010. *Building a Program for Institutional Analysis of Social-Ecological Systems: A Review of Revisions of the SES Framework*. Bloomington, IN: Workshop in Political Theory and Policy Analysis.
- Meinzen-Dick, R. 2007. Beyond panaceas in water institutions. *PNAS* 104(39):15200–15205.
- Ojima, D. S., K. A. Galvin, and B. L. Turner II. 2001. The global impact of land-use change. *Bioscience* 44(5):300–304.

- Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, MA: Cambridge University Press.
- Ostrom, E. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
- Ostrom, E. 2007. A diagnostic approach for going beyond panaceas. *PNAS* 104(39):15181–15187.
- Ostrom, E. 2009. A general framework for analyzing sustainability of social-ecological systems. *Science* 325:419–422.
- Ostrom, E. and M. Cox. 2010. Moving beyond panaceas: An interdisciplinary approach to the study of social-ecological systems. *Environmental Conservation* 37(4):451–463.
- Ostrom, E. and H. Nagendra. 2007. Tenure alone is not sufficient: Monitoring is essential. *Environmental Economics and Policy Studies* 8(3):175–199.
- Ragin, C. C. 2000. *Fuzzy Set Social Science*. Chicago, IL: Chicago University Press.
- Scheffer, M., S. Carpenter, and B. de Young. 2005. Cascading effects of overfishing marine systems. *Trends in Ecology and Evolution* 20(11):579–581.
- Schlager, E. 1990. *Model Specification and Policy Analysis: The Governance of Coastal Fisheries*, Bloomington: Political Science, Indiana University.
- Scott, J. C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.
- Stark, J. and K. Hardy. 2003. Chaos: useful at last? *Science* 301(5837):1192–1193.
- Stephenson, A. G. 1981. Flower and fruit abortion: Proximate causes and ultimate functions. *Annual Review of Ecology and Systematics* 12:253–279.
- Stern, N. H. 2007. *The Economics of Climate Change: The Stern Review*. Cambridge, MA: Cambridge University Press.
- Tang, S. Y. 1992. *Institutions and Collective Action: Self Governance in Irrigation*. San Francisco, CA: ICS Press.
- Tucker, C., J. C. Randolph, and E. J. Castellanos. 2007. Institutions, biophysical factors and history: An integrative analysis of private and common property forests in Guatemala and Honduras. *Human Ecology* 35:259–274.
- Turner II, B. L. and P. Robbins. 2008. Land-change science and political ecology: Similarities, differences, and implications for sustainability science. *Annual Review of Environment and Resources* 33:295–316.
- Walker, B., L. Gunderson, A. Kinzig, C. Folke, S. Carpenter, and L. Schultz. 2006. A handful of heuristics and some propositions for understanding resilience in social-ecological systems. *Ecology and Society* 11(1).
- Wilson, D. S. 2010. *Economics and evolution as different paradigms X: The ultimate-proximate distinction and why it matters* 2010 [cited August 10 2010]. Available from http://scienceblogs.com/evolution/2010/04/economics_and_evolution_as_dif_9.php.
- Young, O. 2002. *The Institutional Dimensions of Environmental Change: Fit, Interplay, and Scale*. Cambridge, MA: MIT Press.