DISCUSSION PAPER

August 2015 ■ RFF DP 15-40

Ecosystem Services Indicators: Improving the Linkage between Biophysical and Economic Analyses

James Boyd, Paul Ringold, Alan Krupnick, Robert J. Johnston, Matthew A. Weber, and Kim Hall

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

Ecosystem Services Indicators: Improving the Linkage between Biophysical and Economic Analyses

James Boyd, Paul Ringold, Alan Krupnick, Robert J. Johnston,
Matthew A. Weber, and Kim Hall

Abstract

For ecosystem services analysis, a key to collaboration between natural and social scientists is the identification and measurement of *linking indicators*: biophysical indicators that facilitate social evaluation, including monetary valuation of ecological changes. As ecosystem service analysts and practitioners better recognize the various ways in which people benefit from ecosystems, natural scientists will be called on to develop, use, and report on metrics and indicators that link to those diverse benefits. The paper develops principles to guide the identification of linking indicators, compares their features with those of more commonly collected ecological measures, and reviews empirical evidence pertinent to their identification, definition, and performance, primarily from the point of view of conducting monetary valuation of ecological outcomes.

Key Words: ecosystem services, ecological indicators, nonmarket valuation

JEL Classification Numbers: Q51, Q57

Contents

Executive Summary **1**

1. Introduction **6**

2. Linking Indicators: Guiding Principles and Theory **10**

Using Production Frameworks to Identify Welfare-Proximal Ecological Outcomes 10

The Importance of Beneficiary-Specific Indicators 12

Why Is It Desirable to Measure Outcomes That Directly Matter? 14

Summary 16

3. The Usefulness of Linking Indicators in Various Types of Environmental Analysis **17**

4. From Concept to Practice: Key Topics for Linking Indicators Research **19**

Research Topics Suggested by Existing Indicator Guidelines 20

Questions Arising from Gap Analysis and Expert Workshops 22

A Benefit-Cost Analysis (BCA) Case Study 25

Significance of the Key Topics for Biophysical Science 28

5. Linking Indicator Literature Review **29**

The Role of Linking Indicators within RP and SP Valuation 31

Literature Review to Answer the Questions Raised in the Key Topics 35

Summary 64

6. Conclusions and Recommendations **65**

References **69**

Ecosystem Services Indicators: Improving the Linkage between Biophysical and Economic Analyses

James Boyd, Paul Ringold, Alan Krupnick, Robert J. Johnston,
Matthew A. Weber, and Kim Hall*

Executive Summary

Public policy increasingly demands insight into the social consequences of environmental policy and drivers of human behaviors that affect the environment. Social consequences can provide potent justifications for environmental protection and management, and human preferences and related behaviors are the key to understanding both the cause of and solutions to most environmental challenges. The firmer the "handshake" between biophysical and social analysis, the better our ability to address those challenges will be. Yet often the indicators measured by biophysical scientists do not correspond with factors relevant to human preferences and behavior.

A key to collaboration between natural and social scientists is the identification and measurement of *linking indicators*: biophysical indicators that facilitate social evaluation, including monetary valuation of ecological changes. Selection and measurement of biophysical

* James Boyd: Resources for the Future, 1616 P Street NW, Washington, DC 20036; boyd@rff.org. Paul Ringold: US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology. Alan Krupnick: Resources for the Future. Robert J. Johnston: George Perkins Marsh Institute, Clark University, Worcester, MA. Matthew A. Weber: US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division, Freshwater Ecology Branch, Corvallis, OR. Kim Hall: US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division, Freshwater Ecology Branch, Corvallis, OR; and Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR.

Acknowledgements: This paper reflects the contributions not only of its authors but also of a number of colleagues. In particular, we are grateful to scientists who reviewed an early version of this document. Their comments led to a wholesale rethinking of our approach, resulting in substantial improvements. They are Dan Auerbach, Edward B. Barbier, Thomas O. Barnwell, Elena M. Bennett, Joe G. Champ, Susan Clayton, Sharon Collinge, John Duffield, Brendan Fisher, Joel Huber, John Loomis, N. LeRoy Poff, George Van Houtven, and Kathleen Wolf. In addition, Ryan Daniels and JoLynn Collins, both of EPA's Office of Air and Radiation, provided efficient and effective management of the contracts (EP-W-08-019 and EP-W-11-029, both with RTI International) supporting this effort. The organizational talents of Mary Barber, the RTI contract manager, helped sustain the effort. We also thank Sharon Barrell of RTI for her efficient editorial support. This work has been supported by EPA's Ecosystem Services Research and its Sustainable and Healthy Communities Research Programs, under Rick Linthurst and Mike Slimak.

indicators have traditionally been driven by natural scientists with relatively little input from the social sciences and (understandably) as a function of measurement cost considerations, regulatory mandates, and ecological theory. For their part, social scientists have traditionally underappreciated the importance of biophysical indicators' definition to the accuracy of their own analyses and have not always employed indicators that link well with outcomes measured and modeled by natural scientists.

The broader deployment of linking indicators—in monitoring programs, policy analysis, public reporting, and environmental accounting—requires that natural and social scientists share a common understanding of these indicators' motivation, features, and practical deployment. This report is geared toward and carries implications for both natural and social scientists. The paper does the following:

- develops a set of principles to guide the further identification of linking indicators;
- describes linking indicators' role in benefit-cost analysis; environmental accounting; and communication of ecological status, trends, and management outcomes;
- compares their features with those of more commonly collected ecological measures; and
- reviews empirical evidence pertinent to the further identification, definition, and performance of linking indicators, primarily from the point of view of conducting monetary valuation of ecological outcomes.

It deserves emphasis that the relative desirability of alternative indicators (in terms of their ability to communicate or contribute to more accurate social welfare evaluation of natural resource outcomes) can be evaluated empirically. Our review of the empirical literature describes several ways in which this can be done. However, even empirically sophisticated practitioners have given relatively little attention to the "which indicators work best" issue. Open questions identified by our review—such as the most appropriate units of account, most appropriate indicators of nonuse benefits, and approaches to aggregation—would benefit greatly from more deliberate empirical examination. Such studies would not only help improve the accuracy and salience of ecosystem services assessments but also lead to greater consensus among practitioners around preferred indicators. Greater consensus would advance the standardization of indicator protocols—a desirable goal because of the need to compare, aggregate, integrate, and transfer monitoring and evaluation results across the nation's ecological and social landscapes.

For any team of natural resource or environmental policy evaluators, we recommend, at a minimum, conceptual development of an ecological and social production framework that describes (1) causal linkages between biophysical outcomes in an ecological system and (2) linkages between biophysical outcomes and social outcomes. Linked production frameworks have several virtues. First, the identification of causal relationships leading to ecological and social outcomes using this framework can facilitate the identification of diverse beneficiaries affected by ecosystem changes. Our review of the empirical literature highlights the significant degree to which the value of natural resource outcomes can depend on their location, type of use or enjoyment, timing, and other beneficiary-specific factors. Ecosystem beneficiaries are diverse and draw value from nature in diverse ways, from consumptive natural resource uses to recreation, aesthetic enjoyment, and ethical and stewardship motivations. This diversity generates a corresponding diversity in linking biophysical measures. Accordingly, an initial broad recommendation is that analysts identify linking indicators in reference to what may be heterogeneous (e.g., demographically, geographically) sets of stakeholders affected by natural resource conditions.

Second, production frameworks permit ecological outcomes (and their indicators) to be differentiated based on the degree to which they directly, versus indirectly, matter to social welfare. Our central hypothesis is that production frameworks help isolate "linking indicators" more proximate to social welfare and that those indicators will (1) be more meaningful and understandable to lay audiences and (2) lead to more accurate and interpretable monetary valuations of ecological outcomes.

Third, production frameworks help identify and organize the full set of models, expertise, and data needed to relate intermediate ("nonlinking" or "distal") ecological outcomes to linking indicators and to resource anagement options, stressors, and conservation actions.

Again, however, our review of the literature suggests that the desirability of linking indicators is empirically understudied and should therefore be treated as a theoretical hypothesis, rather than an unequivocal, generalizable fact. We strongly advocate more explicit empirical testing of this hypothesis. In the meantime, the burden of proof, in our view, lies with advocates of distal outcomes as the most appropriate linking indicators, because, by definition, distal outcomes require one to know (or speculate about) additional ecological production relationships in order to understand the effects of ecological changes on human welfare. In contrast, indicators that are directly proximate to (or directly affect) human welfare are ideally suited as arguments or variables in behavioral or valuation models, because these indicators do not require any further ecological "translation" to identify changes that are directly relevant to people.

Accordingly, our second broad recommendation, assuming the hypothesis is borne out, is that analysts identify via ecological production frameworks the most proximal possible biophysical indicators, given real-world measurement and modeling constraints.

Extending the above concepts, the paper also develops and addresses the following additional empirical questions relevant to putting linking indicators into practice:

- Are ecological contributions to welfare similar enough across beneficiaries that no special targeting of linking indicators to specific beneficiary groups is necessary? If targeting is necessary, how should this be done?
- Do more aggregate, generic indicators perform better than more specific, disaggregated indicators?
- Do indicators that aggregate over multiple categories (such as indices) perform better than indicators that focus on specific ecological features?
- What are the temporal and spatial dimensions of specific ecosystem-beneficiary pairings relevant to the choice of linking indicators?
- Do indictors pertinent to "nonuse" benefits present complications relative to indicators of "use" benefits?

Can generalizations be made about the units (e.g., percentage change versus absolute numbers versus qualitative brackets) used to express a linking indicator?

The paper reviews literature, answers the questions when possible, identifies gaps in our knowledge, and proposes qualitative (e.g., focus group) and quantitative (e.g., survey) research strategies to address those gaps.

As practitioners better recognize the different ways in which people benefit from ecosystems, natural scientists will be called on to develop, use, and report on metrics and indicators that correspond to these diverse uses. And as the research questions identified in this report become more clearly resolved, there will be implications for conducting biophysical measurement programs. Accordingly, it is essential—as we hope this report demonstrates—for natural and social scientists to collaborate on the definition, evaluation, and operationalization of linking indicators.

For those seeking practical guidance on what to measure, we make the observation that the answer is unavoidably a function of the specific issue being evaluated and the financial and technical resources that can be brought to bear. However, for any application seeking

coordinated natural and social science analysis, we recommend that both groups use the relevant ecological production system, its corollary identification of relevant beneficiaries, and (again, subject to empirical justification) emphasis on biophysical outcomes as proximate as practicable to beneficiary experience and behavior.

1. Introduction

This paper emphasizes the importance of indicators designed to link ecological outcomes to social welfare, provides guidance on their key features, and identifies research strategies that can help natural and social scientists define and measure them.

Public policy generates an increasing demand for insight into the social consequences of environmental policy. Social consequences can provide potent justifications for environmental protection and management, and human preferences are the key to understanding both the cause of and solutions to most environmental challenges. The firmer the "handshake" between biophysical and social analysis, the better our ability to address those challenges will be. Our choice of biophysical indicators can facilitate or frustrate that handshake.

Traditionally, ecologists and other natural scientists have taken the lead on the choice of ecological indicators for monitoring, modeling, and mapping. This practice is accepted because they, rather than social scientists, design biophysical measurement and modeling programs. The selection of indicators by natural scientists has been driven by a range of factors, including convenience (what can be easily measured), regulatory mandates, and ecological theory. Often, the choice of indicators reflects a balance among multiple goals and constraints. For example, Hughes and Peck (2008) describe the challenge in selecting measures for national stream monitoring.

Over the years, numerous criteria for ecological indicator selection have been developed (e.g., Cairns et al. 1993; Jackson et al. 2000; Dale and Beyeler 2001; Fisher et al. 2001). These frameworks call for the selection and deployment of indicators that are relevant not only to the natural sciences but also to broader publics. This desire has intensified as decisionmakers increasingly pose policy questions requiring linked natural and social science analysis. Natural and social scientists have responded by deepening their collaborations on the analysis of ecosystem services. Ecosystem services science explicitly requires data, models, and analyses that link ecological conditions to social welfare. Indicators that are valid and meaningful in both the ecological and social realms play a key role in this linkage.

Our focus is on what we term *linking indicators*. These are not the only indicators relevant to ecosystem services analysis. Broader suites of both natural and social indicators are necessary to fully understand ecosystem service provision and value. But linking indicators are crucial because they act as the points of contact between ecological and social systems and analysis.

Biophysical indicators is our general term for biophysical features, quantities, or qualities that can be defined and measured or modeled.

Linking indicators is our term for biophysical indicators that best facilitate social interpretation of ecological conditions and change. These are defined as indicators that measure those things that directly affect people's welfare. We also at times refer to *linking outcomes*: verbal descriptions of what is quantitatively captured by indicators.

Social scientists and policy audiences crave biophysical outcomes that as clearly and directly as possible "matter" (or are relevant) to lay audiences. An understandable natural science perspective is that everything in nature potentially matters, given the complex interrelationships that govern natural systems—a point we do not dispute. But many things that matter scientifically and ecologically do not *clearly* and *directly* matter to lay audiences. The challenge is to link, via ecological analysis, the things that ecologically and scientifically matter to outcomes that resonate with individuals, households, communities, and politicians. Biophysical outcomes that do this will improve the clarity, accuracy, and power of ecological analysis.

The US Environmental Protection Agency's (EPA's) Science Advisory Board (2009) underscored the need for linking indicators by endorsing Schiller et al.'s (2001) recommendation to "develop language that simultaneously fits within both scientists' and nonscientists' different frames of reference, such that resulting indicators [are] at once technically accurate and understandable." However, while providing substantial specific detail on how to design indicators in response to other criteria, these guidelines provide little specific guidance on the kinds of biophysical measures that are most useful for social analysis. Also, an evaluation of the capacity of existing programs to provide ecological information for social analysis identifies numerous shortcomings (Ringold et al. 2012a, forthcoming).

The same can be said of attempts to translate the ecosystem services typology advanced by the Millennium Ecosystem Assessment (MEA) into a practical and coherent system of indicators to facilitate welfare analysis (Wallace 2007; Barbier et al. 2009). The MEA ecosystem services typology (e.g., supporting, provisioning, regulating services) has proved to be an influential and useful communications device. But it is a conceptually muddled typology, ignoring, for example, important distinctions among ecological processes, outcomes, and benefits. The resulting confusion can thwart the identification of linking indicators (Boyd and Banzhaf 2007).

This paper addresses these shortcomings by summarizing and evaluating research on the definition of ecological indicators amenable to social analysis. We also identify a set of

unresolved issues relating to the identification of linking indicators. We set out a research strategy—emphasizing social science methods—for addressing these unresolved issues.

Linked ecological and social analyses, and the ecological information that underpins them, has a broad range of policy applications. Consider two types of analysis particularly relevant to federal resource managers: regional or national ecological assessments and benefit-cost analysis (BCA) of environmental management and regulation. The importance of national ecological assessments is that data at this scale enable strategic analysis of ecological trends, stressors, and associated social impacts. This strategic analysis can lead, over the long term, to priority allocation of resources to develop, support, and apply models to aid management and protection. Data included in regional assessments are also available to construct and evaluate these models. Examples of such assessments are Burgan and Engle (2008) for US coastal waters and EPA's Office of Water (2010) for US rivers and streams. National or regional assessments that facilitate understanding of environment-related social outcomes are likely to not only improve such strategies but also amplify their political relevance.

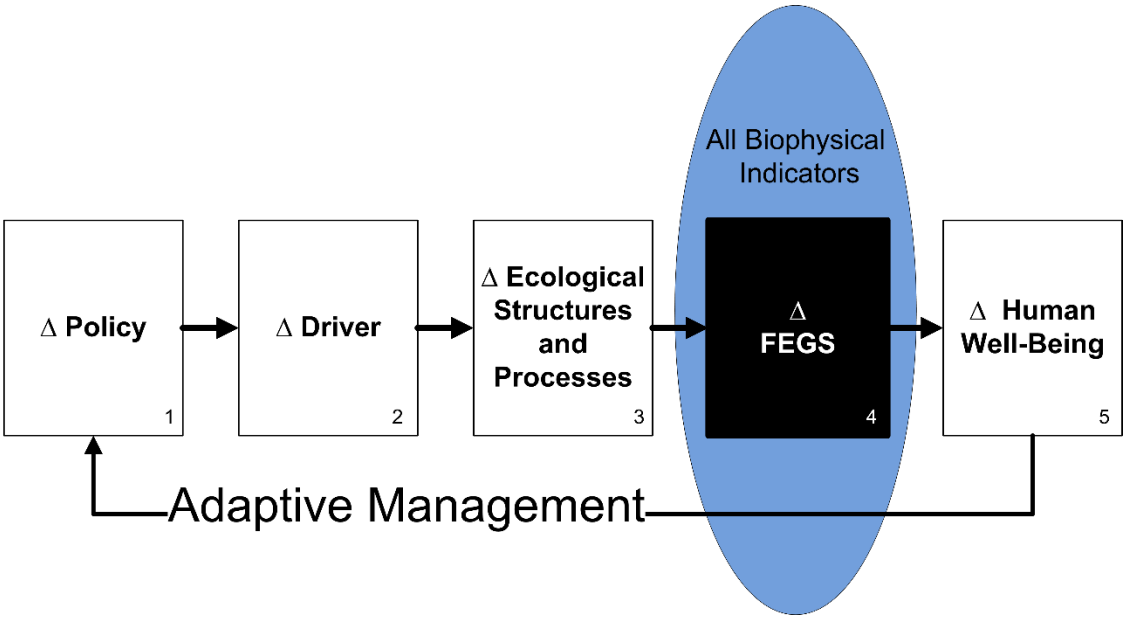
BCA can also be conducted at a national scale but is often applied to more specific federal decisions, such as agency budget priorities, planning guidelines, and regulatory initiatives. BCA requires social and economic assessment of ecological outcomes and thus provides clear motivation for improved linking indicators.

A parsimonious depiction of ecosystem services analysis, and the key role of linking indicators, in these policy contexts is as follows. Various policies—environmental and otherwise—change environmental conditions. We call these changes *drivers*. "Negative" drivers include conversion of natural lands to developed uses. "Positive" drivers include ecological protections that prevent ecological losses or restoration and resource management interventions that improve current conditions. As depicted in Figure 1, these drivers trigger a suite of subsequent ecological changes mediated by ecosystem processes and structures. We refer to this system of drivers and changes as an ecological production system. Understanding such systems is a key aspect of identifying the indicators we seek.

In principle, an indicator can be found for any ecological change. Having a robust portfolio of such indicators is of great scientific importance. In this paper, however, we focus on the subset of these indicators that facilitates social interpretation and analysis. These linking indicators do not by themselves describe welfare changes. The translation of linking indicators into welfare analysis requires additional data (e.g., demographic, market, behavioral, or other human preference data) and analytical techniques (e.g., economic valuation methods). Improved

measurement of linking indicators will enhance these social analyses, but social data and methods to estimate welfare are not our main focus.

Figure 1. Linkages Required in an Ecosystem Services–Based Policy Analysis



Notes: For example, in considering a change in acid rain policy, policy changes in the Clean Air Act could yield measurable changes in facility emissions (the driver) that, via air, water, and biological processes, will yield measurable changes in biophysical outcomes, including chemical water concentrations and fish abundance. Some of these outcomes will be better at facilitating accurate and meaningful social well-being analysis than others.

Section 2 of the paper articulates two broad goals that we argue can help frame the identification of linking indicators. The more obvious of these is the desire for biophysical outcomes that lay audiences can relate clearly to their own well-being. The second goal is to choose indicators that enhance the accuracy of social evaluations. We show how both of these goals are related to the concept of ecological production. Section 3 describes the deployment of linking indicators in various policy applications. Section 4 succinctly summarizes research questions arising from that discussion. Section 5 provides a review of what we know about those research questions from the literature, and Section 6 summarizes our analysis and turns to implementation.

Our hope is that this strategy will facilitate collaboration between natural and social scientists; improved understanding, specification, and measurement of linking indicators; and more accurate and powerful environmental policy analysis.

2. Linking Indicators: Guiding Principles and Theory

As noted in the preface, there have been repeated calls for the definition and measurement of ecological indicators that are relevant to lay audiences and interpretable by social scientists.[1] There has been scant guidance, however, on how to identify such indicators. As demonstrated in this paper, figuring out what directly matters to lay audiences is not as straightforward as it may seem. This paper acts as both a survey of what we know about what matters and a guide to future research. There is no crisp definitional test of an "outcome that directly matters." However, we can apply certain principles in our search for such outcomes.

These principles are derived from "ecological production theory" (Boyd and Banzhaf 2007; Boyd and Krupnick 2013) and other efforts to link ecological outcomes to changes in social welfare (Johnston and Russell 2011; Johnston et al. 2013b). While the principles are defensible theoretically, they embed multiple hypotheses subject to empirical verification. Empirical strategies for testing the hypotheses are taken up in Section 5.

Using Production Frameworks to Identify Welfare-Proximal Ecological Outcomes

One important principle is to describe biophysical outcomes (and their indicators) in terms of an ecological production framework. Production frameworks describe causal relationships (or chains) between inputs and outputs in a system. Ecological production frameworks describe how changes to an ecological feature or quality change other ecological features or qualities.[2]

> An *ecological production framework* depicts causal linkages among biophysical outcomes in an ecological system.

Placing ecological outcomes in a production framework is useful because it allows us to differentiate between inputs and outputs, and to distinguish outputs based on the degree to which they are "distal" versus "proximal" to a given outcome in a chain of causation. Proximal

1 By "lay audiences," we mean any beneficiary of an ecological system. Lay audiences include households, farmers, businesses, community planners, and natural resource managers. They may be technically and ecologically knowledgeable but are distinct from the research community focused on scientific understanding of ecological systems.

2 For example, biological and chemical processes transform water of one quality into water of a different quality. Reproductive, forage, and migratory processes relate biotic and physical conditions to the abundance of species. Food chains convert one form of biomass into another. Wetland processes transform the scale, location, and speed of flood pulses.

outcomes are those closer—causally—to an outcome of interest. For example, nitrogen concentrations in water affect the concentration of algae, which affects the concentration of oxygen in the water, which affects the number and kind of fish that can live in it. Nitrogen levels are a distal input to fish abundance in this chain of causation. Algae and oxygen concentrations are less distal indicators of likely fish abundance.

An ecological production framework is an essential aspect of our approach because it helps us both identify more effective, welfare-relevant indicators and model relationships among indicators.[3]

A second principle is to identify and measure biophysical outcomes that matter as directly as possible to social welfare. What does it mean to "directly matter?" An outcome directly matters when it is valued as an end in itself (Johnston and Russell 2011). An outcome indirectly matters when it affects, or is a necessary input to, a subsequent outcome that affects welfare. For example, fish abundance directly matters to anglers. Oxygen concentrations in water, necessary for fish abundance, indirectly matter to anglers.

Production frameworks, by depicting a system's input-output relationships, can help distinguish between outcomes that matter directly versus indirectly. We will refer to indicators that capture outcomes directly relevant to welfare as proximal indicators. Distal indicators capture outcomes that are not directly relevant to welfare, but that affect proximal outcomes. The interconnectedness of natural systems implies that any proximal indicator will have related distal indicators representing intermediate outcomes in the causal chains that affect the direct outcome.

Scientists working on ecosystem services have increasingly debated and adopted the notion of "final" versus "intermediate" ecological goods and services—a development for which this report's authors are partly responsible (Boyd 2007; Boyd and Banzhaf 2007; Fisher et al. 2008; Fisher and Turner 2008; Kontogianni et al. 2010; Johnston and Russell 2011; Ringold et al. 2013). There is a clear conceptual relationship between that concept and the goals of this paper. The distinction between intermediate and final goods comes from economic accounting, where there is a desire to depict the conventional economy as a complex set of linked production

3 Conservation biology, landscape ecology, restoration ecology, and natural resource management studies already direct a great deal of attention to understanding these kinds of cause-and-effect relationships involving natural systems.

and consumption patterns.[4] Much like an ecological system, an economic system can be depicted as a set of inputs and outputs linked by production relationships. Moreover, economic goods and services can be distinguished based on their proximity or distance to consumers. Bulk iron ore does not directly matter to consumer welfare but is a necessary and valuable input to products like cars and washing machines that do directly matter to consumer welfare. Thus iron ore, an intermediate good, is an outcome distal to welfare, whereas cars and washing machines are final goods proximal to welfare.[5]

The identification of both final goods (in conventional economic terminology) and linking outcomes (in our ecological terminology) motivates a search for outcomes that directly matter to welfare. However, in our experience, the terms *final* and *intermediate* generate distracting and unintended connotations that argue against their usage. For example, some interpret the word *final* to imply a fixed or small set of ecological measures, or they feel the term suggests that other, intermediate ecological outcomes are not valuable or important to measure. Also, there is often a confusing difference between final *ecological* outcomes and final economic goods and services dependent on them. For example, it is not always clear when purely "ecological" production terminates in a "final" good or service, and when "human production" begins (e.g., economists often disagree over whether agricultural products such as corn should be considered an ecosystem service).

For these reasons, we use the term "linking indicators"—biophysical indicators that best facilitate social interpretation of ecological conditions and change—rather than "final" ecosystem goods and services, though the concepts are similar.

The Importance of Beneficiary-Specific Indicators

To this point, we have referred generically to lay audiences and the outcomes that matter to them. But lay audiences are heterogeneous. They are heterogeneous in terms of their

---

4 In order to avoid double counting of value added or social benefits, economic accounts distinguish between intermediate and final goods. For the purposes of gross domestic product accounting, for example, only final goods are tracked and weighted, since the value of inputs is captured in the value of the final product. Counting inputs and outputs in a single economic index double counts the inputs' role in adding value to the final good. See Dernburg and McDougal (1972, 63).

5 As we emphasize later, ecosystem beneficiaries are heterogeneous and interact with natural resources at different points in an ecological system. Using the iron ore analogy, steel manufacturers (as opposed to household consumers) are extremely knowledgeable about the role ore qualities and quantities play in their business operations. So for that class of beneficiaries, bulk ore volume is a very proximate outcome measure.

environmental preferences and the ways in which nature contributes to their welfare. They will also exhibit vast differences in comprehension based on the ways they interact with and experience natural systems. Accordingly, linking indicators can and should be tailored to specific kinds of ecosystem beneficiaries.

For example, soil quality, subsurface water availability, and pollinator populations will directly matter to farmers, whereas they may not directly matter to another type of landowner. The farmers' intimate dependence on, and knowledge of, those outcomes makes these things directly relevant to their welfare. We make this point to emphasize that linking indicators need not, and should not, be "dumbed down" to reflect what directly matters to the least informed or least concerned social audience. Instead, the goal is to tailor linking indicators to the needs and comprehension of more specific types of beneficiaries.

We emphasize that what people perceive as directly mattering to their welfare is a broad and complex empirical question—an issue we return to below in Section 5. Our assertion at this juncture is simply that ecological outcomes can be distinguished via production frameworks by the degree to which they directly (versus indirectly) matter.

Beyond that, we make some initial broad generalizations. Indicators that require technical or scientific jargon are less likely to be linking indicators. It is common in ecology to measure such things as indices of biotic integrity, benthic disturbance, rotifer productivity, hydrogeomorphic features, turbidity, and trophic states. Measures like these usually signify the need to subsequently translate the measure into outcomes more meaningful to nontechnical audiences and proximal to welfare. A benthic disturbance measure is a distal indicator that is meaningless to anyone but those with expertise in stream ecology. Measures of fish and amphibian abundance affected by benthic disturbance are more proximal indicators. Even when an outcome is proximal, jargon can obscure communication and interpretation of the outcome. For example, a water turbidity measure such as "concentration of suspended particles" directly captures something lay audiences may care about—water clarity. But it is difficult for lay audiences to translate the technical (jargon-rich) measure into something understandable (e.g., clear versus unclear water).

Another hypothesis is that environmental outcomes we can see, smell, hear, taste, and touch are more likely to directly matter to our welfare than outcomes we cannot experience directly. The clarity, taste, and odor of water are more likely to directly matter than water quality attributes we cannot detect without scientific instrumentation, such as nitrogen, oxygen, and phosphorus concentrations. Carbon sequestration rates or greenhouse gas (GHG) concentrations

are difficult for us to sense and do not directly matter to welfare. The effect of sequestration and GHGs on precipitation and drought, coastal flooding, and species are easier for us both to experience and to relate directly to our welfare.

Having said that, we also emphasize that direct relevance to welfare does not imply that a linking outcome needs to be physically available to or used by a person. Consider the important case of existence or nonuse values associated with, for example, endangered species. Many people clearly value the existence of certain species even if they will never see, hear, or touch them. The existence of such a species directly matters even without a tangible physical experience of the species.

Why Is It Desirable to Measure Outcomes That Directly Matter?

Why is it desirable to identify proximal indicators of outcomes that directly matter to people? If the goal is to measure ecological outcomes in terms that are relevant to lay audiences, then it is obviously preferable to communicate about what directly matters to them. Distal indicators are proxies for what people directly care about. The problem with proxies is that, even though lay audiences or social scientists may see a causal connection between a distal indicator and their welfare, they may not see the relationship clearly or accurately. Distal indicators introduce confusion, possible speculation, and error when it comes time to interpret the measure in welfare terms. Distal indicators leave unanswered the question of how the proxy relates quantitatively to the outcome that directly matters.

Thus another reason to seek proximal indicators is that they facilitate more accurate valuation of environmental outcomes. Proximal indicators reduce the need for lay audiences and social science researchers to speculate regarding biophysical production relationships. When we ask people to value a distal outcome (or when lay individuals choose behaviors in response to a distal outcome), they are often required to guess at the quantitative relationship between the distal proxy and what directly matters to them. Lay audiences and social scientists lack the knowledge to do this well. For example, if we ask people to relate bald eagle habitat to their welfare, we force them to speculate about the ecological production relationship between habitat and bald eagle abundance. Similarly, if individuals choose a fishing site based on nutrient concentrations in the water, they generally do so based on a speculation regarding relationships between the distal indicator (nutrient concentrations) and things that are more directly relevant to them (fish abundance).

There is a relationship, but its shape and magnitude are unknown. That gap in knowledge (even in such a simple example) makes it nearly impossible to accurately evaluate value or social importance.

To be sure, lay audiences can and will speculate about ecological relationships. For example, those that see a relationship between nitrogen concentrations and their welfare will do so by *qualitatively* intuiting a web of relationships (e.g., nitrogen affects water quality and fish). *Quantitative* speculation (e.g., *how much* nitrogen reduction causes *how much of* a change in water clarity and fish abundance) will almost certainly be wildly inaccurate. If improved accuracy of social evaluation is a goal, it is far better for natural scientists to quantify ecological relationships pertinent to welfare than for lay audiences or social scientists to fill in those gaps on their own.

Similarly, a household or firm owning real estate in a floodplain may understand that wetlands are valuable because they reduce the severity of flood pulses and property damage. But property owners will not understand the shape or magnitude of the hydrological relationship between wetlands and flooding. A distal indicator of wetland land cover is much harder to accurately relate to welfare than are the proximal indicators of flood probabilities, height, speed, and location.

When lay audiences express broad desires for ecological integrity, health, or naturalness, it begs the question of why those things matter to them in more specific terms. Consider an indicator like an "index of ecological health." Such an indicator is a distal proxy for outcomes more proximate to welfare (such as the kinds and abundance of species the ecosystem supports). Unless an index is composed of linking indicators, it is difficult to value or relate to welfare except in very vague terms.

Finally, proximal indicators facilitate the interpretation of ecosystem valuation studies by decisionmakers. Linking indicators are not necessary to detect welfare changes arising from a change in ecological condition. Economists routinely look for relationships between ecological conditions and social welfare. However, when distal measures are related to welfare, this increases the likelihood that more proximal omitted variables are really driving the welfare effect. If so, empirical investigation of the relationship between more proximal omitted variables and welfare may provide clearer, more targeted guidance on how policy can affect welfare via management of those variables.

For example, a study that detects a relationship between nitrogen measures and housing values near a water body (e.g., Jordan et al. 2003; Shields et al. 2008) is clearly measuring a

welfare consequence of an environmental condition. However, it is possible—indeed likely—that nitrogen is simply a proxy for other, more directly welfare-relevant ecological outcomes such as clarity, odor, and fish abundance. Were outcomes such as clarity, odor, and fish abundance available as independent variables, we would be able to test the reliability of nitrogen as a proxy and assess in more detail which ecological elements most strongly drive welfare changes. That latter capability is particularly important for public policy and natural resource management because interventions directed at more specific ecological conditions may be more practical, desirable, or cost-effective.

Summary

We summarize these principles and our hypotheses as follows:

- Ecological production frameworks can distinguish between proximal and distal outcome indicators.
- Linking indicators measure biophysical outcomes that directly matter to social welfare. The more proximal to welfare, the more likely the indicator will be an effective linking indicator.
- Effective linking indicators are those that facilitate communication of ecosystem outcomes to lay audiences, provide more accurate ecosystem valuations, and improve the interpretation of ecosystem valuation studies by decisionmakers.

Our claim that indicators of outcomes closer to welfare will lead to more accurate ecosystem valuations is a theoretically motivated hypothesis in need of empirical testing. Section 5 describes a set of ways in which the hypothesis can be tested.

We hypothesize that ecological production systems are a critical tool for identifying "linking indicators" because they allow us to distinguish between proximal outcomes that matter to welfare directly and distal outcomes that matter indirectly. By definition, linking indicators are more easily understood by lay audiences and minimize the need for lay audiences and social scientists to speculate as they try to relate ecological outcomes to changes in social welfare.

We stress that linking indicators are not the only indicators needed to conduct environmental analysis, nor are they the only things in nature that are valuable. They are simply the indicators most likely to improve the linkage between ecological analysis and public policy, social evaluation, and lay communication.

3. The Usefulness of Linking Indicators in Various Types of Environmental Analysis

A motivation for identification and measurement of linking indicators is that they can play a central role in diverse forms of environmental policy analysis. The linking outcomes they measure can act as the following:

- relatively simple, but powerfully intuitive, measures of an ecosystem's status;
- the goods and services included in "green" economic accounts; and
- the goods and services valued monetarily in economic analysis of a natural resource decision.

Importantly, linking indicators are applicable to and desirable for all of these applications. To illustrate, consider the following concrete, if provisional, linking outcomes: drinkable groundwater volumes, open space acreage, and recreational fish abundance. Note that indicators like "presence or volume of drinkable groundwater," "acres of open space," and "number of native trout and salmon" are countable biophysical features consistent with the principles described in Section 2. For example, drinkable groundwater expresses a biophysical quality in a way that directly matters to many people's welfare. Open space can be tangibly experienced (visually) and can be clearly correlated with behavior (and choices), such as where people like to buy homes or recreate, and is thus likely to directly matter to many people. Fish abundance is a particularly direct biophysical expression of what directly matters to commercial and recreational anglers. Abundance measures also resonate with those of us who care about the continued viability of a threatened or endangered species.[6]

The most direct use of linking indicators is their unadorned presentation as ecological status measures. For example, drinkable groundwater volumes, acres of open space, and fish abundance can simply be reported at regular intervals to describe biophysical trends. As a general rule, current status and trends reports (Council on Environmental Quality 1997; H. John Heinz III Center for Science 2008; US Environmental Protection Agency 2008) include some metrics that correspond to our definition of linking indicators but are mostly composed of metrics that do not pass our litmus tests, because they are scientific and technical measures

[6] To reiterate, these specific indicators may not be ideal. "Acres of open space," for example, is an unnecessarily vague measure of aesthetic qualities—an issue we return to in Section 4.

whose relevance to public concerns and discourse requires further biophysical interpretation and translation (Stoddard et al. 2005; Boyd and Krupnick 2013; Ringold et al. forthcoming).

For environmental accounts, linking indicators can act as the units of account in an index of ecological production or consumption.[7] Such accounts measure, over time, a portfolio of "quantities," considering both individual quantity changes and more aggregate movements. Economic accounts weight the quantities in a way that reflects the quantities' relative importance or value and choose and aggregate quantity measures to avoid both systematic undercounting and double counting. Because linking indicators are defined in relation to production theory, they facilitate the distinctions between intermediate and final goods and thus help avoid double counting.[8] Second, linking indicators by design are easier to weight convincingly. Because the goal of this kind of index is to reveal "valuable increases in production" (or "undesirable losses in production"), accounts need to weight the goods in a way that reflects their relative social importance. Because ecosystem goods and services lack market prices, weights need to be derived, proposed, and debated. Because linking indicators are by design more closely related to social welfare, we hypothesize that their use will reduce technical conflict and likely result in weightings that are more accurate and less controversial than those given to more distal ecological measures.

The goal of benefit-cost analysis is to educate decisionmakers and stakeholders about the costs and benefits of policy or management options and identify actions with the highest net benefit. Linking indicators play a crucial role in this analytical agenda because they represent the biophysical measures that best facilitate subsequent monetary evaluation of costs and benefits. Because they represent the ecological measures most directly associated with human welfare (by definition), they are ideally suited as "things to be valued" within valuation and benefit-cost analysis. As discussed in more detail later, a range of empirical techniques is used to estimate monetary values. All of these methods require as clear as possible a linkage between ecological changes and social outcomes. In fact, ecological outcomes that are not intuitive or clearly related to social welfare can stymie valuation.

---

[7] It should be noted that environmental accounts take a variety of forms. For example, "material accounts" systemically track flows of inputs throughout an economy. And various forms of "economic accounts" exist for the conventional market economy, depending on the goals of measurement (e.g., an emphasis on prices, a particular business sector, or an asset class).

[8] If intermediate and final goods are aggregated at the same time, double counting occurs, because final goods already embody the value of inputs used to produce them.

Linking indicators are desirable, if not necessary, as units of measure for any analysis that seeks to link ecological status or changes to social well-being.

4. From Concept to Practice: Key Topics for Linking Indicators Research

Having established the theoretical and practical motivations for identifying, measuring, and applying linking indicators, we now turn to a set of more specific empirical questions relevant to putting linking indicators into practice. We highlight the eight key topics listed in Box 1.

Box 1. Empirical Research Questions to be Addressed in Identification of Linking Indicators

Key Topics

1. What are the operational guidelines for determining a candidate indicator's success in terms of (1) salience to and comprehension by nonexperts; and (2) capacity to improve the accuracy and reduce the bias of valuation estimates?

2. Do measures that capture outcomes directly relevant to social welfare (proximate indicators) perform better for valuation purposes than more distal indicators?

3. Are preferences similar enough across beneficiaries that no special targeting of linking indicators for specific groups is necessary? How has targeting been done?

4. For any specific linking indicator, do aggregate descriptions (e.g., fish) perform better than more disaggregate descriptions (e.g., trout)? Under what conditions? (This type of aggregation is called in-category aggregation.)

5. Do indicators that aggregate over multiple categories (cross-category indicators, such as ecological health indices) perform better than indicators that focus on specific ecological elements? In the limit, are people able to separate various ecological indicators from very aggregate or holistic measures of ecosystem health?

6. What are the temporal and spatial dimensions of specific ecosystem-beneficiary pairings that matter? Can any generalizations be made about indicator performance along the temporal and spatial dimensions?

7. Does the existence (or nonuse) value context present any specific complications in indicator design relative to the use value context?

8. Can any generalizations be made about the unit of account to express a linking indicator (e.g., acres, percentage change, absolute numbers, qualitative indicators [high, medium, low], icons, pictures)?

In Section 5, we review literature that sheds light on these questions, answer the questions when possible, identify gaps, and propose research strategies to address those gaps.

Topics 1 and 2 address the overarching issue of performance criteria for linking indicators, and our central hypothesis that ecological production frameworks are a critical tool for identifying indicators, respectively.

Topics 3 through 8 address a set of corollary issues. These questions emerge from (1) a comparison of linking indicator criteria with existing ecological indicator guidelines; (2) a set of EPA-sponsored expert workshops held to identify linking indicators relevant to regional and national monitoring programs and subsequent reports on the results of those workshops; and (3) a case study of linking indicators' potential application in benefit-cost analysis of a policy related to acid deposition.

Research Topics Suggested by Existing Indicator Guidelines

Characteristics of ecological indicator specification have been amply discussed from the information producer (usually natural science) perspective (e.g., Cairns et al. 1993; Jackson et al. 2000; Dale and Beyeler 2001; Fisher et al. 2001; Schiller et al. 2001; Niemi and McDonald 2004; Niemeijer and de Groot 2008). Indicators consistent with existing indicator guidelines are necessary, but not sufficient, for the specification of linking indicators. To ensure that the specification of linking indicators reflects both the information producer perspective and the social beneficiary perspective, we examine the 15 guidelines listed in Jackson et al. (2000) and evaluate which, if any, of them require joint specification. These Jackson guidelines are those currently relied on by EPA and are the most detailed set available for ecological monitoring practitioners. We list them in Table 1 and evaluate each for its potential role in specifying linking indicators.

Several of the Jackson guidelines (1, 2, and 15) reinforce or complement the rationale and motivation underlying the development of linking indicators and thus provide no additional requirements for indicator specification. Other guidelines (Phase 2 guidelines 3 through 7, as well as guidelines 8 and 13) address the feasibility of implementation. Measurement feasibility is clearly important but is primarily an issue for natural scientists to address.

Table 1. Characteristics of Ecological Indicators from Jackson et al. (2000) and Their Relevance to Linking Indicators Research Needs

Phase 1: Conceptual Relevance	
Guideline 1: Relevance to the Assessment	Consistent with linking indicator definition and motivating principles
Guideline 2: Relevance to Ecological Function	
Phase 2: Feasibility of Implementation	
Guideline 3: Data Collection Methods	All relate to feasibility and quality of biophysical measurement practices.
Guideline 4: Logistics	
Guideline 5: Information Management	
Guideline 6: Quality Assurance	
Guideline 7: Monetary Costs	
Phase 3: Response Variability	
Guideline 8: Estimation of Measurement Error	An issue for biophysical measurement practice, with a corollary need regarding valuation estimates
Guideline 9: Temporal Variability - Within the Field Season	Time and space need to be defined not only for ecological reasons but also for beneficiary reasons
Guideline 10: Temporal Variability - Across Years	
Guideline 11: Spatial Variability	
Guideline 12: Discriminatory Ability	Consistent with linking indicator definition and motivating principles
Phase 4: Interpretation and Utility	
Guideline 13: Data Quality Objectives	Relates to basic quality of both biophysical and social/economic measurement practices
Guideline 14: Assessment Thresholds	Pertinent to linking indicator definition and presentation
Guideline 15: Linkage to Management Action	Consistent with linking indicator definition and motivating principles

Jackson's Phase 3 guidelines (8 through 12) are more relevant to a social science–related research agenda. They require evaluation of temporal and spatial ecological characteristics. The methodologies provided by the guidelines to support such evaluation focus on managing ecological variability but are silent with regard to how people perceive or benefit from patterns of variability. For example, whereas an ecologically driven definition of streamflow might focus on annual patterns of flow, a description of flow for an irrigator would focus on flow relevant to irrigation requirements. Thus we conclude that beneficiary-relevant specification of temporal and spatial attributes should be considered in developing a research agenda for defining linking indicators. This analysis is particularly important because, while the literature on defining temporal and spatial sampling sufficiency for ecological characterization is extensive, it is sparse for the joint consideration of ecological and social characterization. Key topic 6 reflects this need.

Finally, guideline 14 requires the specification of assessment thresholds. Specification of these thresholds is of importance to "facilitate interpretation of indicator results by the user community." While the document states that the obligation is to partition acceptable from unacceptable conditions using a variety of means, the more fundamental question is, what indicators should be used to express acceptable versus unacceptable ecological conditions so that the threshold is meaningful? How does a measure of, say, fish abundance compare with people's expectation? Should individual measures, such as fish abundance and aesthetic appeal of a location, be aggregated into a single measure, as illustrated in Ringold et al. (2013), or presented as disaggregated measures? Key topic 5 is related to these kinds of questions.

Questions Arising from Gap Analysis and Expert Workshops

EPA's Office of Research and Development has sponsored three recent workshops and analyses focused on linking indicators and current ecological monitoring protocols for stream ecosystems. The projects resulted in a detailed identification and analysis of gaps in the capacity of regional and national monitoring to represent linking indicators for stream ecosystems (Ringold et al. 2012a, 2012b, forthcoming). The projects involved input from monitoring practitioners, natural scientists, and social scientists and explored the degree to which existing monitoring systems developed indicators amenable to social and economic evaluation.

This effort identified gaps that limit our ability to deploy linking indicators at a regional or national scale, including the following:

1. Uncertainty about what quantities, qualities, or conditions are of direct relevance to beneficiaries.
2. Limitations in the ability to translate from currently measured and measurable units to information more directly meaningful to people (proximate indicators) and vice versa. For example, how does the concentration of pathogens translate to making water swimmable or usable for irrigation? How does the desire of a municipal water facility to draw water with less potential to foul intakes translate into water quality measures?
3. Limitations in understanding the temporal and spatial attributes of ecological resources that directly matter to beneficiaries.

Examples of these gaps for two types of beneficiaries are provided in Box 2.

Under the first gap, we identified two major challenges. The first is to specify a set of beneficiaries. This is essential because linking indicators are defined relative to the diverse ways in which people benefit from ecosystems. Systems of total economic value, which partition direct use from indirect use from no-use/passive values (e.g., Bishop et al. 1987; United Nations Environment Programme 2005; Turner et al. 2008), provide an important higher-level classification, but within these broad classes there remain beneficiaries with diverse preferences for ecosystem attributes (Ringold et al. 2009, 2011; Landers and Nahlik 2013). This diversity is reflected in the specification of linking indicators and leads to Key topic 3.

The second major challenge identified was for nonuse beneficiaries. This challenge is evident in the fact that nonuse beneficiaries were addressed in different ways in each of three workshops that attempted to identify linking indicators (Ringold et al. 2009, 2011). Which ecological attributes make a difference to a nonuse beneficiary? Do only biological attributes make a difference, or should we also include other attributes such as chemistry or aesthetic appeal? Within each attribute of relevance, is the valued feature something charismatic (e.g., polar bear presence)? Is the valued feature the deviation of the existing attribute from some expectation for that attribute? If the valued feature is the deviation from some expectation, what is the expectation? Is it for a pristine, least disturbed, historic, or desired condition? Diverse approaches relevant to specifying and evaluating which expectation matters to people have been noted by others (e.g., Stoddard et al. 2006; Clayton and Myers 2009; Jackson et al. 2011). Definition of this expectation is important, because different specifications of the expectation lead to different natural science research needs, assessment thresholds, and resource descriptions. Issues associated with this challenge correspond to key topics 7 and 8.

Box 2. Examples of Indicator Development Using Existing Data: Gaps and Challenges

To document specific gaps in the capacity of national and regional programs to represent the status of linking indicators, we constructed indicators using existing data for two types of beneficiaries: catch-and-release anglers and farm irrigators.

Linking indicators for recreational catch-and-release anglers should include representations of both fish and the places the anglers fish (Hunt 2005). To represent fish, we used a metric reporting on abundance in two families of fish: salmonidae and centrarchidae. We divided this metric into three "quality" classes. We combined this measure of fish quality with a measure of site appeal, also divided into three "quality" classes. We combined the two sets of quality classes into a measure of fishing quality. Details about our methods and a representation of our results are provided in Ringold et al. (2013).

To produce these results, we encountered a number of questions and gaps. For example, we were uncertain about what to measure to describe safety of water contact for this beneficiary. In our illustrative analysis, we did not include this characteristic. We made a number of translational assumptions: which fish taxa of which sizes were of what value to people; how fish abundance leads to fish quality; whether and how site appeal (Peck et al. 2006) can be divided into appeal classes sensible for an angler; and integration of fish quality and site appeal into an index using a specific weighting that, while perhaps reasonable, needs empirical grounding. Our data have temporal and spatial limitations. They arise from a onetime low-flow sample of a defined reach length (40 wetted channel widths for wadeable streams and 100 for larger streams). Whether this sample reflects that collection of fish of relevance to an angler is uncertain. The data are from a probability sample and thus reflect the status of defined population of streams (Olsen and Peck 2008). While this results in the capacity to provide unbiased population estimates, it does not lend itself to supporting the creation of maps that depict the status of all streams. For some purposes, population estimates are sufficient; for others, maps are important. Specification of the temporal and spatial dimensions of an appropriate indicator is a social science question that, if addressed, would lead to an improved specification of the linking indicator. Natural science research (i.e., spatial modeling) would be necessary to address the spatial limitations associated with the density of information, but such research is not necessary to better specify the linking indicator.

Irrigators withdraw large quantities of water from streams (Kenny et al. 2009). Linking indicators should depict not only flows of water sufficient for irrigator needs but also the certainty and timing of streamflow. Irrigators also require water with salinity suitable for the crops and soils under cultivation (e.g., Shannon and Grieve 1998). Finally, the water must have chemical and pathogen levels that leave crops suitable for their intended use. We were able to develop an irrigator-relevant water quality indicator as shown in Ringold et al. (forthcoming) using existing data. To provide estimates of the quantity of water available, we multiplied sample design weights (Olsen and Peck 2008) by average annual streamflow estimates (Hughes et al. 2011).* We categorized salinity generically using guidelines from deHayr and Gordon (2004).

To produce these indicators, we encountered a number of gaps and challenges. We were uncertain about what to measure to describe the safety of water for irrigation arising from factors other than salinity. We were also uncertain about how to represent the certainty of streamflow for irrigators, though historical stream gauge data would be available in many locations to do so. In our analysis, we did not include these characteristics at all. We made one major translational assumption on the linkage between salinity and suitability for use. Our assumption was a generic one, which is sensible for a report of regional or national extent. The wealth of natural science research on crop salinity tolerance** would support great refinement of this translation. Our salinity data have the same temporal and spatial limitations as the data for recreational anglers. In addition, our flow estimates are for annual averages. This is not the quantity of water providing benefits for irrigators.

Box Notes
*The legal capacity to withdraw water is an important constraint on how much water an irrigator can withdraw. However, linking indicators represent quantities or conditions that are available from ecosystems. Issues of access, especially anthropogenic access factors, such as regulations, customs, or infrastructure, are important in determining linking indicator use (or enjoyment or appreciation) but do not determine linking indicator status.
** See, for example, https://www.ars.usda.gov/Services/docs.htm?docid=10135.

Under the second gap, relating to translations, we identified for each beneficiary a number of steps necessary to translate measurements to meaningful information. For example, Secchi disk readings are simple measurements frequently taken in lakes. Hedonic analyses show that water clarity, as quantified by Secchi disk readings, is an attribute related to waterside property values. The question, though, is how Secchi disk readings should be translated into a socially meaningful status report. Heiskary and Walker (1988) illustrate one method. In another example, Ringold and colleagues (2013) developed a fishing quality index based on the intersection between fish abundance and site appeal. This representation assumed that each fish, large or small, of recreational interest or not, contributes equally to human welfare. This assumption is likely not tenable. In a partial translation, Oliveira and his colleagues (2009) developed a Fishery Quality Index that assigns different value weights to different taxa but not to different sizes of fish. Translation of measurements to meaningful information must recognize the variation in preferences or comprehension within a beneficiary class. These considerations lead to key topics 4, 5 and 8.

The third gap, limitations in understanding the temporal and spatial attributes that provide value to people, is the functional equivalent to the question that arises in Table 1 from considering the phase 3 guidelines for ecological indicators (described below) and leads to key topic 6.

A Benefit-Cost Analysis (BCA) Case Study

In general, comparable gap analyses for regulation-oriented benefit-cost analyses do not exist. However, to illustrate gaps and research questions associated with BCA studies, we focus on indicators as deployed in acid deposition studies. While somewhat dated, acid deposition analysis associated with Clean Air Act regulation was particularly extensive, well resourced, and

geared toward both ecological assessment and economic valuation. It is therefore illustrative of the state of practice in terms of BCA.[9]

Studies on aquatic ecosystem benefits arising from reduced atmospheric deposition have focused on two distinct beneficiaries and the ecological features associated with them. One approach has focused on recreational fishing (e.g., Englin et al. 1991; Macmillan and Ferrier 1994). These analyses, consistent with Notes: , link changes in control policies to changes in emissions, changes in deposition, changes in aquatic ecosystem chemistry, and ultimately changes in the presence or abundance of specific fish taxa of recreational interest. For anglers, "presence or abundance of specific fish taxa" is an approximation of the relevant linking indicator.[10] There are, however, important questions about the precise form of the linking indicator. Should it be the number of fish, the number of fish over a certain size, the presence of fish, or something else?[11] The final step in this linkage, the linkage between changes in lake chemistry and fish (i.e., between boxes 3 and 4 in Notes:) can be quantified because of the existence of models linking aquatic ecosystem chemistry to fish presence or abundance (Baker 1991; Baker and Christensen 1991). This suggests two main criteria. The first, consistent with key topics 4, 5, and 8 is the precise specification of resources that directly matter to people and how to translate a measure into meaningful information. The second is the existence of the set of articulated models that quantify the intertwined relationships between changes in policy and changes in linking indicators. These models are of the utmost importance in benefit-cost

[9] In 1980, President Carter signed the Energy Security Act of 1980. One portion of that act, Title VII, became known as the Acid Precipitation Act. That act created an Interagency Task Force on Acid Precipitation. During the next decade, the task force managed expenditures of about $1 billion (in 2014$) in a coordinated research and assessment program. Fully half of these resources were allocated toward ecological research. After the Clean Air Act Amendments of 1990 mandated a market-based control program for acid rain, the research and assessment activities continued, albeit with a much lower level of support. Given the large, sustained, integrated research on acid deposition, consideration of valuation studies for this system should give us insight into valuation in the best case. Despite this investment, attempts at developing benefit analyses find significant limitations. These limitations illustrate the need for an approach focused on both linking indicators and some of the limitations in our capacity to meet this need.

[10] While anglers respond both to fish and to site appeal (Hunt 2005), acid precipitation valuation studies have focused only on changes in fish and not on changes to site appeal, even though changes in acid deposition or emissions of acidic precipitation precursors could affect both.

[11] Some valuation studies are built around catch per unit fishing effort (CPUE). We would argue that while useful, such measures are not the linking indicator. CPUE combines an ecosystem property (i.e., the number of fish) with technological and human skill factors that result in CPUE. Because linking indicators should more strictly reflect ecosystem properties (see Figure 1 in the preface and the surrounding discussion), CPUE cannot be a linking indicator.

analysis. However, the absence and shortcomings of such models are richly noted elsewhere (e.g., Heal et al. 2004; Daily et al. 2009; US EPA Science Advisory Board 2009; Banzhaf and Boyd 2012). Further, this strategy focuses on the specification of linking indicators; as such, questions with regard to the models that predict them are outside the scope of this strategy.

While one set of acid deposition benefits studies has examined the direct use benefits experienced by anglers, nonuse beneficiaries also gain from changes in sensitive lakes attributable to reductions in acid deposition. However, here the absence of an indicator of ecosystem status (and models linking changes in deposition to that indicator) is a barrier to valuation. Three studies illustrate this. Chestnut and Mills (2005) conclude that "quantitative assessment remains problematic due to a lack of units of measure to gauge changes in the quality and quantity of ecosystem services and a lack of dose–response relationships to indicate how quality and quantity may change as a function of changes in pollution exposures." In a similar, but earlier, analysis, Burtraw and his colleagues (1997, 35) note that "while non-use values of ecosystem health are expected to be large, there is no characterization of ecosystem changes associated with Title IV or of a valuation framework for assessing benefits from improvements in ecological indicators, especially given the temporal aspects of ecological dynamics." In a more recent study, Banzhaf and his colleagues (2006) state, "Until now, all of these abatement initiatives have taken place in the absence of economic estimates of the total benefits that would result from improvements to the park's ecosystem. In part, this mismatch is explained by the large health benefits that independently justify most policies that reduce acid rain precursors. *But this mismatch has resulted primarily from inadequate information on the ecological effects of changes in emissions and deposition and an inadequate link between the ecological science and social science necessary to enable economic valuation of the benefits of emission reductions*" (emphasis added). The absence of a nonuse indicator, highlighted in these acid deposition studies, leads to key topics 3 and 7.

Notably, while the specification of beneficiaries is one of the key elements of identifying linking indicators, the way this is approached in benefit-cost analysis may differ from the manner in which it is addressed in status and trends monitoring. In the former case, an analyst is better able to study a system and then identify and engage those benefiting from it. In effect, the identification of beneficiaries and linking indicators relevant to them can be tailored to the specific system and changes in question. In contrast, the design of a regional or national status and trends report must capture a broader, more generic set of outcomes, benefits, and beneficiaries. While it may be desirable to delineate beneficiaries comprehensively, as suggested in Ringold et al. (2013), it may be more practical to identify a generic set of diverse beneficiaries

with competing interests that is not comprehensive. In this case, the specification of the linking indicator may need to be more general than in a benefit-cost analysis. This observation shapes the ways in which key topic 3 should be addressed.

Significance of the Key Topics for Biophysical Science

These key topics can be addressed by conducting social science research. However, each has important implications for conducting biophysical science to deliver linking indicators. As the first two key topics are addressed, the usefulness of the linking indicator concept will be evaluated, and guidance and methods to discriminate between "better" and "worse" linking indicators will be provided. This will help natural scientists make defensible decisions about the most useful metrics and indicators to measure.

As we better recognize the different ways in which people benefit from ecosystems, natural scientists will be called on to develop, use, and report on metrics and indicators that correspond to these diverse uses. For example, national reports on the status of aquatic ecosystems are currently designed to report on ecosystems in terms of biotic integrity (the maintenance and restoration of which is a key objective of the Clean Water Act). If people value aquatic ecosystems for multiple values, then effective reports might be constructed around multiple indicators reflecting diverse values. Any definition of which values are most important can come only from the recognition that there are multiple beneficiaries and then by analysis of which are most important, as well as specification of the criteria used to operationalize "most important." Depending on the construction of those indicators, this may require changes in the way data are collected. For example, indices of biotic integrity are constructed without regard to the potential for biota to foul water intakes. However, fouling of water intakes can be a significant problem for power plants and water supply systems, with annual costs estimated in the hundreds of millions of dollars (Nakano and Strayer 2014) and even higher (Isom 1986). Thus national surveys that would seek to report on aquatic resources important to these beneficiaries would need to collect information on fouling potential.

If more aggregate descriptions of resources (e.g., fish present) are more salient than less aggregate descriptions (e.g., good-sized trout), then there are consequences for the ways in which resources are assessed and managed. If aggregate descriptions are found to be more salient, that will, in general, provide support for simpler forms of data collection, reporting, and modeling. In contrast, if less aggregate descriptions are more salient, then not only is there more reason to provide more monitoring detail, but there also may be reasons to allocate management resources

differently. For example, the proportion of western stream miles with "good-sized trout" present is about one-fifth of the proportion that has fish "present."

If more composite indices are more salient than indicators that focus on specific ecological indicators, then the composition and ways in which components are combined may have implications for what data are collected or modeled and with what precision. For example, if a fishing quality index is much more heavily influenced by site appeal than by the abundance of fish, then effort should be allocated to ensure the collection and adequacy of the site appeal metrics.

A rich literature exists on how to sample to represent ecological resources in the face of their temporal and spatial variability (e.g., Holme and McIntyre 1971; Green 1979; Hughes et al. 2002; Mueller-Dombois and Ellenberg 2002). The literature on the temporal and spatial scales of ecological features that constitute final ecosystem goods and services is sparse (as described in the next section). To the extent that the temporal and spatial dimensions that people experience or perceive vary from those required to define ecological features, then methods that accommodate this dual requirement must be developed.

Finally, nonuse beneficiaries are important (Freeman et al. 2014). As the features constituting final ecosystem goods and services for these beneficiaries are defined, examining monitoring and modeling protocols would be essential to ensure that they can represent these features.

5. Linking Indicator Literature Review

The purpose of this section is to review the economic valuation literature and set out a gap-filling research agenda for linking indicators, as defined in prior sections. Our foundational hypothesis is that ecological production frameworks are a critical part of identifying linking indicators that provide the most direct link between environmental changes and human preferences and welfare. We would furthermore argue that resultant linking indicators should be considered a primary means of tracking policy or program outcomes and should also be used as the primary basis for nonmarket valuation of environmental changes. We hypothesize that valuation analyses that do not use linking indicators risk bias and increased uncertainty in value estimates due to the indirect and often ambiguous relationships between nonlinking indicators and human welfare.

This section is written primarily for an audience of both natural scientists and economists, although the findings are also relevant to other social sciences that study human-environment

interactions. For economists, the goal is to characterize the types of biophysical indicators that are best suited for use within models of human behavior and welfare (i.e., value), with a particular emphasis on revealed preference (RP) and stated preference (SP) valuation. We also describe a research agenda to identify and test alternative linking indicators and to permit judgments about the best indicators. For natural scientists, the goal is to characterize the types of indicators that are most directly relevant and meaningful to the public (i.e., that are directly relevant to human welfare) so that these can be compared with indicators that are currently used. This information can also be used to inform programs that model, monitor, or otherwise measure biophysical indicators so that these programs might target indicators more directly relevant to human welfare and behavior. This latter goal builds on prior gap analysis. This includes Ringold et al. (forthcoming), who compare linking indicators derived from an expert workshop (Ringold et al. 2009) with current regional and national practice. In doing so, they identify a research agenda for altering currently used indicators to better align with those meaningful to the public.

While the general concept of linking indicators is (arguably) straightforward, the details in specific empirical applications may be less so. Given these complexities, the goal of this section is to both clarify the meaning and role of linking indicators within valuation and discuss questions that must be addressed to enable the identification, measurement, and use of linking indicators in both the natural and social sciences.[12]

The section begins with a brief discussion of the role of linking indicators within the context of the two methodologies for valuing ecological or environmental outcomes within the economic literature: RP and SP approaches. We then continue to address the eight key topics. These key topics were designed to clarify the definition and valuation of linking indicators. This discussion is followed by a review of the economics literature (primarily associated with valuation), summarizing applicable research methods and findings pertinent to the questions

[12] For example, consider the definition of linking indicators relevant to improvements in fish habitat within lakes used by recreational anglers. Even for a single beneficiary group (anglers), isolation of an unambiguous set of linking indicators may not be obvious. Are the most relevant linking indicators very narrow (e.g., abundance of individual species and size classes most valued for harvest), narrow (e.g., abundance of individual species and size classes legal for harvest), or broad (e.g., fish abundance)? Are some linking indicators a combination of other indicators and not necessarily all biophysical (e.g., catch per unit effort [CPUE], which depends on both natural conditions and human/technological capital)? Do relevant linking indicators vary across different types of beneficiaries (e.g., shore versus boat anglers)? Moreover, even when the general outcome valued by individuals (e.g., water appearance) is known, the most suitable linking indicators to characterize those outcomes (e.g., clarity, turbidity, color) are not always clear.

raised in key topics 1 and 2. This is supplemented with a partial review of closely related findings in other research literatures. We conclude with a discussion of implications for future research. It is our hope that future work can more comprehensively elucidate the treatment and relevance of the linking indicators within other areas of social science, such as anthropology, marketing, psychology, and sociology.

The Role of Linking Indicators within RP and SP Valuation

Linking indicators have a different place in the RP and SP literatures. In the RP literature, the goal is often to model behavior as a function of resource quality and quantity variations. An example is the discrete choice travel cost model, in which individuals' choices to visit different recreational sites are modeled as a function of variables that include the cost of travel to the site, as well as the characteristics of substitute sites. Within such models, observable behavior is generally modeled as a function of readily available measures of ecosystem output or status. RP valuation studies are tied directly to ecosystem attributes, such as various beach characteristics, which are then related to the behavioral metric (e.g., beach visit frequency) and, ultimately, to a monetary metric of value. Among the challenges of indicator selection for these models is the fact that the factors directly relevant to behavior are not always purely biophysical. An example is recreational trout fishing, where average catch per unit effort (CPUE) is a standard metric used to predict frequency of fishing and, ultimately, how much fishermen value this experience. In this case, the metric combines effort—a behavioral variable—and catch, which is also behavioral but related to an ecosystem output, the size of the targeted fish population (e.g., trout).

In nearly all cases, researchers conducting RP research rely on biophysical or other indicators that are *already available*, regardless of whether these indicators are ideal for the task or even whether they are purely biophysical indicators. This reliance on existing indicators provides less flexibility when choosing indicators and leads to frequent applications in which proxy variables or indirectly related metrics are used to characterize the relevant factors of interest. In such cases, the ability of RP researchers to use appropriate linking indicators depends on the prior collection of data on these indicators by natural scientists and existing monitoring programs. Of course, there is nothing to prevent RP researchers from collecting data on ecosystem attributes as part of valuation research. However, these data are rarely collected because of constraints on time, money, and expertise within research projects, a common lack of biophysical scientists collaborating within these projects, and a lack of understanding about which attributes are truly linking indicators for the behaviors of interest.

While RP researchers often give careful attention to the ecological indicators used within their models, in many cases a statistically significant correlation to behavior is used as justification that the "right" indicator has been selected, even when a superior indicator might be found via a thorough search of other available indicators. This common tactic also overlooks the potential errors-in-variables problem that can occur if the indicator used within an RP model is an inaccurate measure of the true linking indicator (i.e., the variable that influences behavior). That is, there may be a poor fit between the indicators of ecosystem quality used within RP models and the drivers of individual behavior related to these ecosystems, leading to unknown biases in model results. The ultimate consequence could be that environmental and resource managers manage the wrong indicator, draw perverse conclusions, or take perverse management actions.[13]

In contrast to RP valuation, which relies on observable conditions and behavior, SP valuation relies on responses to hypothetical scenarios generated by researchers. SP surveys ask questions that allow respondents to state their value for specified, hypothetical but realistic changes in environmental conditions, compared with current conditions (or a business-as-usual scenario that would occur in the absence of any change). Although these scenarios should be grounded in actual conditions and realistic environmental outcomes under proposed policy changes, the SP researcher has more or less full control over the information provided to respondents. Within this valuation context, there is more flexibility to identify and use linking indicators best suited to communicate the environmental changes to be valued.

However, even within SP valuation, there are limits to indicators that can be used. A common example is limits on indicators created by the need to characterize conditions under the no-policy-change status quo. To enable the elicitation of valid values, it is important for SP

13 A good example of this problem is the vast RP literature linking mortality risks for workers in various jobs or sectors to the wages they receive. These statistical analyses provide a very important number—the value of a statistical life (VSL)—which EPA and other agencies have used in benefit-cost analyses of lifesaving pollution regulations. The literature is vast because of the abundance of labor market data that make these analyses possible. However, these analyses poorly match the context for which the VSL is actually used. Air and water pollution affect mortality risks to specific population groups—the very young, the very old, and the very sick— more than others. None of these groups are in the labor force (except perhaps some of the very sick). And the types of risks faced in the labor force are primarily auto accident risks, which are familiar and not dreaded, in contrast to risks from air pollution (such as dying from emphysema or cancer). Analogously, RP methodologies have been developed around available data but may not be well suited for revealing preferences for environmental improvements.

studies to communicate environmental conditions accurately under both the status quo (with no changes) and hypothetical alternatives. Because SP researchers must quantify actual environmental conditions under the status quo, they must (1) rely on indicator data that natural scientists have in hand to quantify current conditions; (2) collect and analyze indicator data as part of the project itself; or (3) develop proxy attribute descriptors that fit with respondent mental models (e.g., as learned about in focus groups or cognitive interviews). For example, to quantify willingness to pay (WTP) for an improvement in water clarity from current levels, one must have information on current water clarity against which to compare proposed hypothetical alternatives.[14] This type of analysis typically requires preexisting measures of the ecological indicator in question, the ability to measure indicators of water clarity as part of the project, or a transformation of existing scientific metrics into a simpler form that is more readily understood by respondents. As most SP projects do not include the resources or expertise to collect new biophysical data, researchers must often use preexisting indicators to characterize the current conditions, or simplifications of these indicators. This is similar to the constraint that often faces RP valuation.

However, unlike the case with most RP valuation, SP researchers often conduct focus groups and pretest surveys that use different biophysical indicators to quantify environmental changes. These steps are an important component of SP survey design (Mitchell and Carson 1989; Champ et al. 2003; Bateman et al. 2002; Johnston et al. 1995; Louviere et al. 2000; Dillman et al. 2014). This can provide information that may be used to select the linking indicators most closely related to respondents' values (Zhao et al. 2013). As with RP studies, significance in statistical models is not sufficient alone to determine whether a meaningful or appropriate linking indicator has been used. For example, many past SP studies have used poorly defined, inaccurate, or immeasurable ecological indicators within survey scenarios yet still generated statistically significant results (Schultz et al. 2012).

More specifically, SP studies often develop estimates of the WTP for resource improvements based on scenarios that are not characterized using well-defined indicators (Schultz et al. 2012). For example, survey designers often respond to the need to quantify

[14] The ability to connect WTP estimates to proposed policy measures also requires a capacity to link these measures to changes in stressors and ultimately to changes in the linking indicators included within the SP scenarios. Linking these measures to changes in stressors is typically accomplished via ecological modeling. Without these linkages, WTP is interpreted as being conditional on a particular environmental outcome being achieved.

environmental conditions with limited data by using characterizations of environmental change with no clear linkage to measurable indicators (Johnston et al. 2012). This situation arises, at least in part, when the data collected by natural scientists for their research projects do not correspond to linking indicators required for economic applications, or when indicator data are simply unavailable. However, it is also due to a lack of familiarity among many economists with methods used and data generated by ecologists and environmental scientists (Swallow 1996; Simpson 1998).

Given these challenges, this section has relevance for both RP and SP valuation, along with the identification of associated environmental and resource features to be managed. For RP researchers, it can help identify the types of environmental indicators that would be more ideal candidates for inclusion within behavioral models. For SP researchers, it can help identify the types of environmental indicators that are meaningful from the perspective of natural sciences and to the public. As noted previously (Johnston et al. 2012; Schultz et al. 2012; Boyd and Krupnick 2013), the use of such indicators within SP scenarios is necessary to avoid potential biases related to respondent speculation and lack of correspondence between survey responses and measurable environmental outcomes.

Grounded in this capacity to improve the validity of economic valuation, one of the primary benefits of the linking indicators perspective is that it can help prevent misguided management actions based on incorrectly specified models (i.e., managing for the wrong things). Incorrectly specified RP or SP models (ones that do not include the relevant linking indicators) can lead to misguided conclusions regarding specific outcomes valued by the public. For example, an SP study might report a positive WTP for reducing nutrients in a water body, when it is possible that what really influenced survey answers was respondents' speculation that reduced nutrients would be linked to improved recreational fishing and birding (e.g. Johnston et al. 2013). In this case, SP results could lead managers to allocate funds to environmental projects that might not, in fact, enhance public welfare. Similarly, managing a fishery to maximize short-run CPUE (an indicator that combines natural and human elements) could have negative long-run natural or social consequences if these short-run CPUE improvements are obtained using technological means that diminish stock or habitat quality in the long run.

The section proceeds as follows. We first give a general description of the key topics presented in Box 1. These questions must be addressed when seeking to identify linking indicators suitable for policy analysis (focusing primarily on valuation). The questions focus on both the identification and performance of linking indicators for specific applications (e.g., valuation studies) and for the types of linking indicators that might be collected regularly (e.g., as

part of monitoring programs) to support policy analysis. We then discuss the literature relevant to each question, again focusing primarily on economics and valuation but also introducing key findings from other areas of the literature, as appropriate. We conclude with a discussion of implications and next steps for research.

Literature Review to Answer the Questions Raised in the Key Topics

The set of key topics developed in Section 4 (see Box 1) guides this critical evaluation of linking indicators and their role in valuation. To address these questions, we thoroughly searched the economics literature (with much less effort in related literatures), seeking relevant evidence on the objective definition and use of linking indicators in the context of SP and RP valuation. We performed searches using key words pertaining to the question topics in databases that include published research, book chapters, and working paper series, such as PsychINFO, EconLit, Research Papers in Economics Social Science Research Network, Environmental Valuation Research Inventory, Web of Science, and Google Scholar. We paid particular attention to how relevant studies could be classified with respect to each question in terms of empirical evidence, assumptions, and underlying conclusions and their generalizability. Once we identified a set of key papers, we conducted searches within citing articles to explore additional relevant research. We also searched for working papers by authors with multiple relevant studies. Finally, we ran searches of the databases again to ensure that these searches were picking up the key papers that had already been identified. Following are the questions raised in the eight key topics along with the literature relevant to each one.

1. *What are the operational guidelines for determining a candidate indicator's success in terms of (1) salience to and comprehension by non-experts; and (2) capacity to improve the accuracy and reduce the bias of valuation estimates?*

This question reflects perhaps the most fundamental concern when choosing indicators for use in either RP or SP studies: How are researchers to know when their ecosystem indicators are actually "working" or serving their intended purpose within the valuation model? Or, alternatively, how can researchers determine whether better indicators may be available? In an RP context, if particular attributes are found to be significant predictors of behavior, they are generally assumed to be working, or providing accurate and unbiased estimates relating changes in the environment to changes in behavior, value, or both. Often these indicators are not strictly biophysical; examples include indicators of beach congestion, the presence of a boardwalk near a beach, and CPUE. But sometimes they are (or are related to) ecosystem attributes; examples include measures of water clarity, absence of jellyfish, and fish population size. As noted above,

however, statistical significance alone is an insufficient measure of indicator performance, because poorly designed proxy variables can also lead to statistical significance in RP models. In the terminology of previous sections, the idea is to find indicators that both are appropriate for valuation (i.e., are the direct or proximal biophysical drivers of behavior) and link back to the ecological processes that support it.

In general, beyond statistical significance of regression coefficients, the RP literature provides relatively little systematic attention to the methods used to select indicators. One area of RP research in which more attention has been afforded is the literature addressing the choice of perceived versus actual measures of environmental quality in RP or RP/SP models (e.g., Adamowicz et al. 1997; Poor et al. 2001). For example, hunters may prefer to visit sites where the *perceived* wildlife abundance is higher, even if this perception does not comport with objective measures of abundance. Results of these analyses have been mixed, with some studies finding that subjective measures outperform objective ones (e.g., Adamowicz et al. 1997) and others finding the opposite result (Poor et al. 2001). These analyses are again based on the performance of alternative variables in statistical models. Similar analyses have been conducted within other literatures, including assessments of the performance of objective geographic information system measures versus subjective measures of landscape aesthetics (Frank et al. 2013).

The SP literature has devoted more attention than the RP literature to the selection of indicators, because this is a critical step in survey design (Johnston et al. 2012) and a key part of content validity in particular (Bateman et al. 2002; Bishop 2003). The process of attribute selection within discrete choice experiments requires that researchers choose a small number of attributes to communicate the outcomes of the policy scenarios considered by respondents (Bennett and Adamowicz 2001; Bateman et al. 2002). Where these policy scenarios involve environmental or ecological outcomes, this process unavoidably requires the selection of a set of ecological indicators “best suited” to communicate these outcomes (Johnston et al. 2012; Schultz et al. 2012). While the evidence provided by this SP survey design work is sometimes indirect, it has provided insight into the effectiveness of different types of ecological information within SP valuation and the methods that may be used to evaluate this effectiveness.

The first and most direct set of methods used in SP research to evaluate indicators is applied during survey design and typically includes focus groups and interviews (Desvousges and Smith 1988; Johnston et al. 1995; Kaplowitz et al. 2004; Powe 2007). Zhao et al. (2013), for example, discuss a formal process for selecting indicators within SP survey design that combines conceptual model development, focus group and interview results, and biophysical data. Similar

process steps are outlined by Van Houtven et al. (2014) and Johnston et al. (1995, 2012), the latter of whom used ethnographic methods within focus groups for SP survey design. Verbal protocols or "think aloud analysis" can be used within focus groups or interviews to further illuminate respondents' understanding of different ecological indicators (Schkade and Payne 1994). The use of focus groups and interviews in this way (e.g., to select attributes that communicate ecological effects in SP scenarios) is a standard element of SP design (Bateman et al. 2002; Mitchell and Carson 1989; Champ et al. 2003; Louviere et al. 2000). Related methods can include pilot surveys of policymakers, experts, or the public or the use of Delphi approaches (Bennett and Adamowicz 2001, 48).

As in RP analysis, evaluation of statistical properties and statistical significance within SP models can provide evidence regarding whether an indicator is relevant to respondents' preferences. Hundreds of related and relevant analyses have been conducted as part of the broader debates over SP validity conducted over the past three decades (Arrow et al. 1993; Hanemann 2006; Carson 2012; Kling et al. 2012). However, as with RP analysis, the relevance of this work for indicator selection is indirect, and the related insights can be misleading. For example, many aspects of survey design beyond the definition of ecological indicators can affect responses and hence the apparent link between an indicator and WTP.

Although there are many ways after a survey has been implemented to evaluate the role, impact, and relevance of ecological information, a small set of methods has emerged as primary. Perhaps the most common of these is the scope test, both internal and external,[15] which evaluates the sensitivity of WTP estimates to changes in the quantity or quality of an environmental good, typically quantified using some form of indicator (Giraud et al. 1999; Bishop 2003; Heberlein et al. 2005). Although these tests (with a few exceptions) have generally supported the validity of SP methods, they have not been used extensively to help choose (or provide insight into the choice of) ecological indicators for use within SP scenarios. An exception is Zhao et al. (2013), who assess whether scope sensitivity (e.g., WTP for ecological improvements) is consistent across alternative indicators used to communicate the effects of river restoration on migratory fish passage. This study finds that different indicators can sometimes be used with equal success

[15] An external scope test compares WTP estimates across different groups of respondents who are shown different levels (quantities or qualities) of an environmental improvement. A positive and significant finding of scope sensitivity implies that WTP is positively related to the quantity or quality of an environmental change. An internal scope test is similar to an external test but compares WTP estimates by the same group of respondents when shown different levels of improvement.

to quantify valued ecological changes in SP valuation, as long as they are chosen using an appropriate process.

Design and analysis of debriefing questions can also be used to gain insight into the appropriateness of the indicators chosen in an SP survey. Some studies ask respondents directly if they felt that an attribute chosen was the "right" one. More frequently, respondents are asked if they thought about indicators of ecosystem change that were broader than defined (e.g., abundance of all species improving rather than abundance of specific species) or if they thought the categories were inseparable. This is often conducted as part of evaluations of potential embedding effects, where embedding refers to the issue of whether alternative goods represent different quantities of the same underlying good (i.e., are nested subsets) or instead represent related but not formally nested goods (Carson and Mitchell 1995). Debriefing questions can also be used to test respondents' knowledge to evaluate whether they understood the presentation of particular indicators. For example, the survey of Johnston et al. (2012) includes questions to evaluate whether respondents understood the uncertainty associated with indicators of fish passage restoration. Nonresponse rate is another way of gaining insight into indicator appropriateness (Arrow et al. 1993); a high nonresponse rate to the valuation question could imply that the linking indicator is poorly defined. Of course, there are many other reasons for nonresponse.

Other general types of SP validity tests can be used to evaluate the indicators used to quantify ecological change. However, as above, the insight from these approaches is often indirect. Moreover, unless carefully conducted using split-sample tests in which the *only* variation across samples is the ecological indicators used to quantify policy effects, the effects of indicator selection can be difficult to distinguish from the effects of other survey design aspects. Examples include a variety of construct and convergent validity tests. The former compare SP results with those expected on the basis of theory. The latter compare SP results with those generated using other valuation methods or compare the results of multiple SP analyses whose results are expected to be similar. For example, meta-analyses of prior valuation studies could be used to evaluate systematic patterns in results associated with using different types of indicators within survey design. Examples include Johnston et al. (2005a) and Van Houtven et al. (2007), who evaluated patterns in SP WTP estimates for water quality improvements, in part as a function of the type of changes communicated in the original surveys. However, like many of the statistical tests identified above, a limitation of such evaluations is that they evaluate only whether WTP differs as a function of the type of indicators used; they cannot by themselves identify which indicators are superior.

Applicable research to gauge indicator performance also exists outside the valuation literature. Note that the two ways of judging performance listed in the key topic span qualitative and quantitative criteria; both types of research are needed to gauge the success of a given indicator. Indeed, qualitative and quantitative research in general each have known weaknesses when used alone, thus the literature on "mixed methods" approaches (e.g., Weber and Ringold 2012; Creswell 2014). Qualitative methods widely utilized in the social sciences, such as anthropology, psychology, sociology, marketing, and health care, yield well-developed protocols for collecting and analyzing verbal and text data (e.g., Morgan and Krueger 1998; Fern 2001; Patton 2002; Rubin and Rubin 2005; Bernard 2011; Creswell 2013; Miles et al. 2014). These protocols include exploring the "in vivo" language in which people describe a given issue, and they have been applied to identify ecological indicators relevant to the lay public as a research goal in and of itself.

For example, Schiller et al. (2001) employed focus groups and interviews to evaluate indicators best suited to communicate environmental conditions to the public. Ringold et al. (2009, 2011) used expert opinion to identify indicators for streams, lakes, and estuaries for a variety of beneficiaries. Weber and Ringold (2014, 2015a) found recurring indicators for rivers based on focus groups and interviews with sociodemographically diverse members of the general public. Focus groups and interviews offer the unique capacity for tailored follow-up questions to ensure that the indicators discussed are considered important in and of themselves, rather than predictive of something else ultimately more relevant. In all of the above studies, a limited set of proximate indicators were successfully distinguished from supporting intermediate or distal indicators.

Another technique normally employed with small samples is Q-methodology, which is specifically designed to study subjectivity (Stephenson 1993). Unlike focus groups and interview methods, which tend to search for themes that emerge from the participants, a participant in a Q-methodology study would rank predefined qualitative variables based on some criteria (e.g., direct relevance to them). Yet another qualitative technique is analytical hierarchy process (AHP) models (e.g., Duke and Aull-Hyde 2002). For example, using AHP, one could compare the relative weight or importance given to different ecological indicators when comparing across policy outcomes; this type of analysis can provide nonstatistical insight into which indicators are viewed as most relevant.

In terms of applicable quantitative methods, content analysis offers systematic techniques to quantitatively analyze qualitative text to explore a research question (Bernard 2011; Krippendorff 2013). Texts can be collected either as primary data or as sampled secondary data.

Researchers then extract information through automated or manual coding of texts. For example, the frequency of various ecological features in the texts can be recorded and compared. Weber et al. (2014) use this approach to analyze blogs and newspaper articles for indicators manifest within the topic areas of rivers, streams, and creeks. While many other examples of content analysis appear in the environmental literature, we are not aware of any other examples focused on documenting or selecting indicators for use in subsequent analysis.

One could also apply various survey methodologies to gauge indicator salience and comprehension, similar to the debriefing questions mentioned in SP work. A well-vetted survey method oriented to qualitative variables is best-worst scaling. Within this approach, survey respondents repeatedly select a subset of text variables as "best" or "worst" from short lists; in sum, these selections can be translated into importance weights for each variable (Finn and Louviere 1992; Flynn et al. 2007). To our knowledge, such approaches have not yet been applied to evaluate ecological indicator preferences, but these methods would seem to offer a powerful means to test hypotheses about linking indicators developed through preceding qualitative research.

In summary, many methods can be used to evaluate the extent to which different indicators are salient and comprehensible to different groups and simultaneously to minimize potential speculation about how the indicator relates to human welfare. However, as of yet, no unitary, generalizable, consensus approach exists for selecting the "right" set of linking indicators for use within RP or SP valuation. Within RP analysis, nearly all methods emphasize the statistical significance of associated model coefficients rather than direct tests of indicator salience. Within SP analysis, similar tests can be matched with the results of qualitative research methods such as focus groups, interviews, and verbal protocols to inform indicator selection. Valuation practitioners would seem to have an opportunity to borrow more deeply and broadly from qualitative and quantitative methods outside the traditional valuation toolbox. Among the challenges for the linking indicators framework is the ambiguity that often accompanies indicator selection for valuation purposes (Zhao et al. 2013) and the need for a consensus set of approaches.[16]

16 Issues such as this are not unique to environmental valuation. For example, Reynolds and Rochon (1991) discuss related concerns from the perspective of advertising research.

2. *Do measures that capture outcomes directly relevant to social welfare (proximate indicators) perform better for valuation purposes than more distal indicators?*

This question arises from our underlying hypothesis that ecological production frameworks are a useful tool for isolating indicators that matter most, a hypothesis suggested by Boyd and Krupnick (2013). At issue are the causal linkages through which biophysical changes influence people (Blamey et al. 2002). An answer of yes to this question would vastly simplify the task of finding linking indicators, because all other possible indicators (e.g., those related to ecosystem processes) or those related to "intermediate" outputs (e.g., nutrient loadings as opposed to algal blooms) could be eliminated from consideration, leaving the focus on proximate indicators (what some term "final" ecological outcomes or final ecosystem services)(Brown et al. 2007; Johnston and Russell 2011; Boyd and Krupnick 2013; Johnston et al. 2013b), also referred to in this paper as linking indicators.

In general, RP analyses rely solely on secondary data—that is, modeling users' observed behavior (e.g., trips to a recreational site) based on a set of independent variables (often including measurable ecological or environmental indicators). As discussed earlier, there is less attention to indicator selection than is common within SP studies. Hence, the manifestation of this issue is primarily statistical. However, as a rule, RP studies usually endeavor to use *directly experienced variables* within behavioral models, unless such variables are unavailable. While not an overt usage of ecological production theory, this does promote a focus on a category of ecological outcomes that are more likely to be directly relevant to behavior, ceteris paribus. That is, a variable that is directly experienced by a person can serve as a direct behavioral stimulus—in contrast to nonexperienced variables, which cannot influence behavior directly (because the individual cannot experience or perceive them). Hence, all else equal, we argue that directly experienced variables are more likely to be directly welfare relevant and thus serve as proximate linking indicators.

Where RP studies do not use directly experienced variables, the variables that are used are usually acknowledged as statistical proxies. For example, Chang et al. (2014) use measurements of pathogen and dissolved oxygen (DO) levels in water in lieu of directly experienced variables of water odor and fish and wildlife populations, which they hypothesize as more directly relevant to behavior. In such cases, the question is whether these proxies have suitable statistical properties to prevent bias or minimize confidence intervals (e.g., due to measurement errors or omitted variables) in resulting regression models. In this case, the core issue and challenge are no different from those in any econometric model in which proxy variables are used to substitute for unmeasured variables of interest.

Within SP analysis, there is considerable indirect evidence that proximate indicators outperform similar distal indicators, again even if ecological production theory has not been explicitly invoked. The only direct test of this proposition, however, is that presented in the currently unpublished work of Johnston et al. (2014). Within the SP literature, there is significant evidence endorsing the idea that knowledge and familiarity17 are important elements for well-defined preferences (Brown and Siegler 1993; Johnston et al. 1995; Bateman et al. 2008). As noted by Bateman et al. (2008, 129), "Experimental and stated preference evidence suggests that when unfamiliar goods are presented in previously unencountered hypothetical market institutions (such as often occurs in CV surveys) resulting initial valuations are liable to be based upon poorly formed preferences. In such situations, the 'constructed preference' literature would suggest that such responses are prone to be influenced by a variety of choice heuristics and framing effects resulting in apparently anomalous preferences." Similar points are made by Hutchinson (1995) and many others. Cameron and Englin (1997) found that respondents more experienced with the environmental resources being valued have both different WTP estimates and smaller conditional variances on these estimates. Lack of experience has also been postulated to help explain at least some of the difference between RP and SP WTP estimates (List and Gallet 2001). Additionally, the literature notes that information can be provided as a substitute for experience. But the literature is mixed on whether more information actually reduces the variance of WTP (Hanley and Munro 1992; McClelland and Schulze 1992; Czajkowski et al. 2015). None of this research, however, addresses the specific issue of utilizing ecological production frameworks as an aid to identifying which ecological indicators may be more directly related to respondents' welfare.

To our knowledge, the Johnston et al. (2014) study is the sole quantitative analysis designed to directly compare the performance of input and process (distal) indicators versus proximate indicators. This analysis begins with a theoretical model characterizing conditions under which respondents can express valid preferences over distal indicators. The authors then tested theoretical results and hypotheses using an application of choice experiments to migratory fish restoration. Both theoretical and empirical results of this analysis suggest that SP models cannot, in the general case, estimate valid preferences for distal input or process indicators.

[17] We use the term *familiarity* to include experience with a good or attribute but also to include goods and attributes respondents may have thought about, such as endangered species.

However, these results do not provide generalizable guidance on the usefulness of ecological production frameworks.

Indirect evidence is provided by Milon and Scrogin (2006), who tested for a difference in WTP between an ecosystem input and output. The sample was split so that half the respondents received a survey that characterized improvements to the Everglades through, in our language, "dual commodities" (water levels and timing), and the other half received a survey that characterized improvements through a particular "output" of the system (species populations). Improvements were defined in terms of percentage similarity of water levels and timing or population levels of three wildlife species groups to historic predrainage conditions. The authors found a significant difference: the former treatment yielded lower WTP. Unfortunately, the interpretation of these findings with regard to theory is not straightforward; there is no unambiguous theoretical expectation regarding whether a distal indicator should be associated with higher or lower WTP estimates, ceteris paribus. Nonetheless, these results do show that the choice of proximate versus distal indicators for valuation influences resulting WTP estimates. To our knowledge, there have been no controlled tests evaluating the effect of proximate versus distal attributes (or indicators) on the variance of SP WTP estimates.

A practical issue associated with administering a SP study applies here. Assume the unlikely circumstance that ecological production models are well understood surrounding a given ecological issue, and that baseline and scenario of change impacts on linking indicators are available. As the list of linking indicators grows as impacted by a given program or stressor, expectations on SP experimental design needs increase dramatically, as do expectations on respondents to cognitively process and make trade-offs across multiple outcomes (Louviere et al. 2000). Given this problem, SP practitioners will sometimes present respondents with scenarios defined by admittedly distal indicators, believing the omitted variables associated with the inevitable speculation to be less important than the omitted variables associated with failing to include all cascading outcomes of a given environmental policy change. We are not aware of formal quantitative tests gauging the relative importance of such sources of omitted variables bias. A related issue is that people sometimes infer impacts on linking indicators outside the intended scope of a SP survey, even if the survey has been carefully developed to include only linking indicators (pretesting by Weber in support of EPA ICR No. 2484.01).

Even more at the heart of the hypothesis is whether people can or do think of ecosystems in ecological production terms. This issue may be best approached through qualitative research. Indeed, some qualitative evidence associated with SP valuation suggests that participants may view environmental issues holistically (Brouwer et al. 1999), and other authors critical of

valuation in general note that participants may not see nature as easily decomposable (Vatn and Bromley 1994). We have also had focus group experiences in which respondents prefer aggregated or seemingly all-encompassing measures such as ecosystem "health" or "integrity" in addition to or instead of disaggregated indicators (e.g., Johnston et al. 2011). Aggregated metrics can certainly serve as linking indicators if they are actually the most proximate metrics possible to social welfare. However, verification is needed that such indicators are neither proxies for other, more relevant outcomes nor a means of avoiding the cognitive challenge of making decisions at a disaggregated level. After all, every aggregate indicator is composed of individual pieces, and there is no single way to build an index. The ambiguities and challenges of aggregate indicators are well established in the ecological literature (e.g., Suter 1993).

An approach that could be employed more often in focus groups is to provide information on ecological production to see whether this allows disaggregated indicators to be better identified. Even if detailed scientific knowledge of ecological production relationships is lacking, the focus group moderator can ask whether a given indicator is important to participants in and of itself, or whether participants *believe* it is predictive of something else more directly related to their welfare. In this vein, Weber and Ringold (2015b) report on their experience utilizing ecological production theory to identify linking indicators for rivers. They describe numerous barriers participants faced identifying linking indicators and provide moderation strategies (i.e., follow-up questions) to address these difficulties. As other qualitative research has found, as mentioned above, one of the important issues was that participants did not always appreciate the mutability of the environment—that is, the range of potential environment conditions. Group meetings were helpful for envisioning possibilities and coming to a point of being able to express preferences. Difficulty thinking about disaggregated environmental changes is potentially related to an extant belief the public holds in the "balance of nature," which other authors have noted and criticized as naïve (Ladle and Gillson 2009; Hovardas and Korfiatis 2011). If nature is truly in balance, managing a finite set of discrete attributes by themselves appears incomplete. While it has been possible in at least some studies (Weber and Ringold 2014, 2015a) to overcome a reticence by focus group and interview participants to compartmentalize nature, more such studies to test this success are needed.

In summary, the combined quantitative evidence from RP and SP studies—while mostly indirect—suggests that using indicators that are more proximate to human welfare within an ecological production framework may improve accuracy and reduce bias. This improve occurs, we surmise, as a result of modeling changes in behavior and value as a function of the variables that directly influence these changes, rather than variables that influence these changes only

indirectly through a chain of ecological production. For example, within SP valuation, the use of more proximate indicators can reduce cognitive errors, confusion, speculation, and scenario rejection. However, additional quantitative work is required to establish this hypothesis with less ambiguity. As for evidence from qualitative studies, again there is some evidence that ecological production frameworks are a useful heuristic to guide identification of indicators directly relevant to human welfare. However, few studies directly testing this idea with human subjects are available.

3. *Are preferences similar enough across beneficiaries that no special targeting of linking indicators for specific groups is necessary? How has targeting been done?*

This pair of questions primarily addresses issues of degree in targeting linking indicator definitions to specific beneficiary groups and trade-offs. It is clear that different linking indicators can be relevant to different user and nonuser groups (e.g., in the context of SP survey development, as discussed above). What is less clear is when differences between groups are of sufficient magnitude and relevance to warrant the identification of these distinct groups ex ante, given the time and resource costs of developing and administering distinct SP survey versions to these groups. For RP approaches, it may even be impossible to take into account different beneficiary groups, because the approaches themselves often target specific groups (e.g., anglers).

On the existence of distinctive preferences, Johnston and Russell (2011), for example, argue that in an SP context, indicators need to be matched to beneficiary groups because what is considered a final ecosystem good will differ by beneficiary (see also Ringold et al. 2009, 2011, 2013). Therefore, empirical valuation of ecosystem services should start with an assessment of beneficiary groups. They note that "final ecosystem services are beneficiary dependent; this dependence is central to any effort to categorize services" (Johnston and Russell 2011). For example, water clarity may be a relevant linking indicator—with Secchi depth an associated linking indicator—for residents who live in a local area and value clearer water. However, water clarity may not be a relevant linking indicator for commercial anglers, who are interested primarily in the capacity of an ecosystem to produce harvestable fish. Hence, the relevant set of linking indicators will depend on the set of beneficiary groups for a particular set of ecological processes. Daw et al. (2011) make a parallel conceptual argument, focusing on the effects of ecosystem services on the welfare of the world's poorest populations and highlighting the heterogeneity in the ways that poor populations benefit from ecosystem services. Such distinctions in linking indicators across beneficiary groups are also important if we are designing

modeling monitoring and mapping programs, since each linking indicator may become associated with a distinct management goal.

Relevant beneficiary groups may be based on economic factors (who actually gains or loses as a result of ecosystem changes) as well as distributional factors (whose "benefits count" for a given policy analysis) (Loomis 2011). Spatial characteristics of beneficiaries (e.g., where they live relative to ecological changes) may also be directly relevant. For example, residents generally value ecological changes "close to home" differently than identical changes at a greater distance (Banzhaf et al. 2006; Johnston and Duke 2009). In the SP literature, heterogeneity between groups is often discussed in the context of users and nonusers of particular resources (Hanley et al. 2003; Hoehn et al. 2003; Johnston et al. 2005b). However, similar variations can occur across any number of different groups. As above, the relationship between ecological outcomes and beneficiary groups is an empirical question and cannot be characterized using natural or social science data alone. These relationships may also vary across different sites at which ecological research is conducted. As argued by Hoehn et al. (2003), the record on knowledge and indicator definition is mixed but largely untested.

Hence, the primary question for valuation studies is not *whether* differences across different beneficiary groups can justify different SP survey designs (with different linking indicators), but *when* differences between beneficiaries are sufficient to justify such action and *how* these situations should be identified and addressed within an SP survey. To a large extent in the literature, the same methods as those discussed under key topics 1 and 2 apply here, including the use of focus groups and interviews to identify whether different groups express preferences of sufficient heterogeneity to warrant the use of different indicators within survey design. That is, the same methods used to identify relevant linking indicators are used to identify whether the same linking indicators apply to different groups of beneficiaries.

Among the challenges in this area of research is a lack of direct empirical evidence for the relevance of different *indicators* to different beneficiary groups and whether/how this is relevant to survey design. As in key topic 2, above, there is substantial indirect evidence that values for different ecological outcomes are beneficiary specific, but this by itself is not direct evidence that indicators need to be different for different beneficiary groups. For example, the SP literature contains hundreds of publications demonstrating that WTP varies according to demographic factors such as education and income. Other studies demonstrate variation in SP WTP according to observable factors such as length of residence in a community (e.g., Johnston et al. 2003), experience with particular activities such as recreational angling (e.g., Cameron and Englin 1997), whether a person lives in or out of an affected area (e.g., Brouwer et al. 2010),

distance between a person's home and affected areas (e.g., Bateman et al. 2006), a person's knowledge regarding an ecological outcome (e.g., Heberlein et al. 2005), user versus nonuser status (Johnston et al. 2005a), and many others. Recent evidence also suggests that preferences for indicators of drinking water quality may differ systematically across male and female members of the same household (Rungie et al. 2014).

Yet these and similar evaluations are typically based on split-sample comparisons of the way that *different groups* respond to the *same survey scenarios*, with the *same indicators*. More to the point, Van Houtven et al. (2014) found differences in the salience of aggregated versus disaggregated indicators, related at least somewhat to experience (as a proxy for differences across beneficiaries). For example, they report that focus group participants distinguished between different groups of fish (what are generally termed game fish and rough fish) and that nonanglers understood that improved water quality supported greater species diversity, which can be described by using more disaggregate descriptors. Nevertheless, to our knowledge, there has been no systematic, quantitative evaluation of the performance of different ecological indicators across different respondent groups, when used to evaluate outcomes of the same policy. Different ecological indicators could serve as the most relevant linking indicators for different beneficiary groups for a number of reasons, such as that (1) different beneficiary groups may benefit from different types of ecological outcomes, each with its own set of linking indicators; and (2) different beneficiary groups may understand indicators differently, even for the same valued outcomes. Evaluations of this type will be required to draw nonspeculative conclusions about whether and how to vary indicators when designing SP surveys for different beneficiary groups.

4. *For any specific linking indicator, do aggregate descriptions (e.g., fish) perform better than more disaggregate descriptions (e.g., trout)? Under what conditions? (This type of aggregation is called in-category aggregation.)*

While key topic 3 applied to beneficiaries, this set of questions applies to the indicators themselves. These questions apply not only to SP studies, but also to RP studies in which disaggregate data on ecosystem endpoints are available and can be aggregated or kept disaggregated for analysis. An obvious example is recreational fisheries data; here an RP study could use catch or abundance indicators at many different levels of aggregation to predict angler behavior (e.g., all game fish, specific types of game fish, and so on).

As with the three key topics discussed above, most of the evidence on this question in the RP literature is related to statistical tests evaluating whether models are improved through the use of aggregate versus disaggregate indicators. This evidence is found in both primary studies

and meta-analyses that combine data across multiple primary studies. Primary studies have often found distinguishable effects associated with relatively disaggregated indicators. For example, various hedonic analyses have found distinguishable impacts on property values associated with multiple disaggregated indicators of water quality, including levels of chlorophyll-a, nitrogen (N), phosphorus, cyanobacteria, and *Escherichia coli* (Egan et al. 2009; Netusil et al. 2014), although interestingly, none of these appear to qualify as linking indicators, since people are unlikely to be able to interpret the impact on their welfare from such technical metrics. That they are included in the models suggests that they serve as proxies for features that are relevant to individuals. Results such as these suggest that different effects can be measured for different (disaggregated) indicators.

In contrast, meta-analyses often suggest that more aggregated indicators and categories better explain variations across different primary studies. The meta-analysis of recreational fishing values by Johnston et al. (2006) found that model properties (including statistical fit and variable significance) were improved when targeted fish species were aggregated into composite groups. For example, they use an aggregated "panfish" category to identify WTP values associated with small (generally low-value) freshwater fish such as catfish and perch, rather than include separate indicators for each individual species. Moeltner and Rosenberger (2008), moreover, describe methods that can be used within meta-analyses to determine whether models should use more or less aggregated data (e.g., to determine whether observations from running-water fishing should be aggregated with observations from still-water fishing). They also found evidence to support analyses over aggregated categories.

Evidence from the SP literature is also mixed but suggests that the relevance of disaggregated categories is often related to respondent experience. For example, survey development of Johnston et al. (2002) led to the use of six disaggregated subcategories of wetland birds in a survey directed to salt marsh restoration experts but only a single aggregate category in a survey directed to the public. As noted under key topic 3, Van Houtven et al. (2014) similarly found differences in the salience of aggregated versus disaggregated indicators reporting that focus group participants distinguished between different groups of fish (game fish and rough fish) and that nonanglers understood that improved water quality supported greater species diversity, which can be described using more disaggregate descriptors. They ultimately used several categories of fish in their SP survey but did not disaggregate them by specific species. In addition, EPA (Federal Register 2013) found that nonanglers did not express distinct preferences for different subgroups of fish in the Chesapeake Bay. Anglers in focus groups distinguished between different fish subgroups but were able to answer questions with effects

described over all fish. These results were used to develop an aggregated indicator (i.e., fish) for the subsequent survey. Johnston et al. (2012) also found that non-expert respondents distinguished between effects on commercial and noncommercial fish populations but did not distinguish between effects on specific commercial or recreational species.

Many examples are found throughout the SP literature in which different types of indicator aggregations are used. These examples suggest that there is no "one size fits all" approach to indicator aggregation and that the appropriate degree of aggregation depends on respondent preferences that may vary markedly across different valuation contexts and beneficiary groups (related to the discussion of beneficiaries above). For example, SP surveys addressing policy effects on fish have expressed policy effects over individual species (e.g., Puget Sound chinook salmon; Lew and Wallmo 2011), well-defined subgroups with distinct ecological characteristics (e.g., small freshwater diadromous fish such as river herring; Johnston et al. 2013a), or highly aggregated groups of all species. All of these efforts were grounded in focus groups and other survey development to help inform the most suitable presentation of policy effects. Studies addressing effects on birds, however, often use more aggregated categories. For example, SP analyses addressing wetlands restoration often present aggregated indicators of effects on bird life (e.g., Johnston 2002; Morrison et al. 2002), at most disaggregating across broad categories (e.g., songbirds versus wading birds; Milon and Scrogin 2006). Banzhaf et al. (2006) provide information about specific threatened tree and bird species in their survey about reducing acid deposition in the Adirondacks, but the linking indicator to be valued is forest cover (acreage) and bird populations.

Complicating the choice of indicators for SP survey design is the fact that the indicator aggregations that are most useful to the public (often nonexperts) may not always comport with classifications used by natural scientists. For example, survey development of Christie et al. (2006) supports the differentiation of "familiar" and "unfamiliar" wildlife, whereas Johnston et al. (2002) distinguish effects of suburban land use on "large mammals (e.g., deer, fox)" versus "small mammals (e.g., mouse, squirrel)," among other species groups. Similar evidence is found in the ongoing work of several of the authors of this document. For example, focus group participants selected from a population at large in arid Arizona were concerned about "tall trees" but not the kind of tree, and in general, an "affinity for particular species of flora or fauna was rare" (Weber and Ringold 2015a). Indeed, this lack of emphasis on particular species extended even to endangered species, where the fact that they were endangered mattered, but not the type of species. Schiller et al. (2001) similarly report a difference in the aggregation, presentation, and specificity in indicators most useful to experts versus nonexperts.

Zhao et al. (2013), note that decisions over indicator aggregation and selection are also indirectly related to the extensive literature on embedding and part-whole effects in SP analysis. As mentioned above, embedding refers to the issue of whether two or more goods represent different quantities of the same underlying good (i.e., they are nested subsets) or instead represent related but not formally nested goods (Carson and Mitchell 1995). Part-whole evaluations compare the sum of WTP for the "parts" to WTP for the sum of the parts combined (the "whole"). Although methodological difficulties with "adding up" tests such as this are well known (Heberlein et al. 2005; Hanemann 2006; Carson 2012), results of these tests reveal that the sum of WTP for disaggregated categories are often not equivalent to (and generally larger than) WTP for a more aggregated whole. Although the implications of the extensive embedding literature for the choice of the *most appropriate* level of aggregation are unclear, similar psychological concepts influence considerations of embedding and indicator design. For example, effects on certain classes or groups of ecological effects (e.g., species groups) may be more salient to people than effects on more disaggregated categories (e.g., individual species). Respondent experience is also relevant. For example, Heberlein et al. (2005) found that when respondents know more about the part and have more experience with the part (compared with the whole), they are likely to assign higher economic values to the part than the whole.[18]

Related insights may be drawn from attitude theory; this would consider the relevance of "affective scope"—liking the whole more than the part—and "cognitive scope," knowing more and thinking more about the whole than the part (Heberlein et al. 2005). Where both affective and cognitive scope are present, we anticipate that respondents would tend to express a higher WTP (or behavioral intention) for more aggregated or holistic categories of environmental change. Conversely, a respondent may show something akin to "negative affective scope"—liking the part more than the whole—or "negative cognitive scope," knowing more and thinking

[18] Heberlein et al. performed a within-subjects experiment where each respondent valued both the part and the whole for four environmental goods: water quality, biodiversity, spear fishing, and wolf populations. Individual respondents were then assigned to one of three economic scope categories for each of the goods: (1) those who showed positive scope (i.e., valued the whole more than the part); (2) those who showed negative scope (valuing the part more than the whole); and (3) those who showed no scope by giving the same value (possibly zero) for both part and whole. The authors found that water quality and spear fishing passed conventional economic scope tests, and respondents also showed cognitive, affective, and behavioral scope sensitivity with these goods. However, wolves and biodiversity did not pass the economic scope test, and respondents showed negative affective and cognitive scope toward these goods (valuing the part more than the whole).

more about the part than the whole. In this case, attitude-behavior theory suggests that a higher WTP (or behavioral intention) would be expressed for more disaggregated categories or indicators. These suppositions, however, have yet to be verified empirically.

Taken together, the combined evidence from the SP literature seems to support a number of stylized facts related to indicator selection, all of which are subject to challenge and require subsequent testing. First, as discussed above, greater respondent expertise often leads to the salience of more disaggregated categories, although this hypothesis requires additional testing to be established with any degree of confidence. If true, this would suggest that the most effective linking indicators may be different for different beneficiary groups—here related to the level of indicator aggregation. Iconized or rare species also may support the use of disaggregated (e.g., species-level) indicators (e.g., Jacobsen et al. 2008; Lew and Wallmo 2011). In contrast, low experience or nonuse goods may be most effectively communicated using more aggregate indicators that are more easily understood by nonexperts. In all cases, however, careful survey development and pretesting are required to evaluate the level of indicator aggregation that is most suitable in any particular circumstance. Insights provided by the survey development process—including focus groups and interviews—are more relevant than these or other stylized facts. As described in the review of key topic 1, we are aware of few qualitative studies involving human subjects oriented solely to identifying ecological indicators, but it is worth noting that these have tended to produce indicators aggregated within categories (e.g., game fish, rather than trout) (Schiller 2001; Weber and Ringold 2014, 2015a).[19]

In summary, the existing literature suggests that the relevance of aggregated versus disaggregated indicators for behavioral modeling and valuation is case-specific, but that some generalizable patterns may apply. To the extent that patterns have been verified directly or indirectly, findings suggest that disaggregated indicators tend to be more relevant to scientific experts, users, and those with more extensive experience with the resource in question. In contrast, aggregated indicators tend to be more relevant to laypersons, nonusers, and those with little or no experience.

19 Potential complicating factors in testing for the appropriate aggregation level are the effects of subtle language differences in surveys on WTP. Such differences could include longer information treatments for disaggregated treatments and changes in question order. Bateman et al. (2002) finds that the more information for a given attribute, the higher the WTP. See Cai et al. (2011) for discussion of ordering effects on WTP.

5. *Do indicators that aggregate over multiple categories (cross-category indicators, such as ecological health indices) perform better than indicators that focus on specific ecological elements? In the limit, are people able to separate various ecological indicators from very aggregate or holistic measures of ecosystem health?*

The discussion of key topic 4 considered aggregated versus disaggregated indicators within the same general class (e.g., aggregated versus disaggregated indicators of fish abundance). For key topic 5, in contrast, we look across a more diverse set of indicators to ask whether these can be meaningfully consolidated into aggregated or multimetric indicators, such as "ecosystem health" or indexes such as the Index of Biotic Integrity (IBI) (Karr 1981, 1991; for use in valuation, see Johnston et al. 2011). There is debate in the ecological literature regarding the use and validity of these types of indicators for various purposes (Niemi and McDonald 2004). The argument for such indicators is that people often view and value certain ecosystem components as part of an integrated whole. Such indicators often "play an important management role by helping characterize ecological condition" (Niemi and McDonald 2004, 102). Arguments against such indicators include potential sensitivity of index scores to choices involving included subindices and index structure, as well as skepticism over the scientific interpretability of holistic categories such as "ecological integrity" and the definition of ecological reference conditions (Stoddard et al. 2006; Johnston et al. 2011). Compared with indicators of specific ecosystem features, multimetric indicators are also less directly related to management objectives given in terms of specific ecosystem features.

Despite significant attention to these issues (and a mature literature) in the natural sciences (e.g. Niemi and McDonald 2004), the empirical SP literature on this issue is thin and based on work by a small number of authors. A study by Johnston et al. (2011) valuing restoration of migratory fish passage in the Pawtuxet watershed of Rhode Island suggests that respondents value outcomes communicated using both aggregated indicators and indicators that focus on specific ecological elements. Respondents gave nontrivial WTP for both direct indicators (acres accessible to migratory fish) and a biological index (IBI), suggesting that studies that focus only on the direct effects of a policy may overlook some of the ways in which a policy may affect utility. Moreover, counting on respondents to use direct impacts as a proxy for other relevant ecological outcomes will likely lead to respondent speculation and biased WTP estimates. This supposition was tested by Johnston et al. (2013b). This study compared responses to two SP surveys—one that includes a multimetric index of biotic integrity as an indicator of ecosystem condition and another that is an otherwise identical survey that omits this indicator. Results suggest that respondents speculated regarding policy effects on ecosystem condition

when the multimetric indicator was omitted and that this speculation led to upwardly inflated WTP estimates for specific policy impacts (e.g., effects on fish populations).

Similarly, Krupnick found, in his focus group experience, that most people have a basic understanding of the "web of life" and have some difficulty disaggregating this web into component parts for valuation purposes. Weber and Ringold (2015a) found something similar, saying people spoke of the environment as a "monolithic concept" and that it was initially challenging for them to disaggregate. These findings are consistent with earlier results of Schiller et al. (2001), who note that "the best approach [when communicating indicators to the public] was to describe the kinds of information that various combinations of indicators could provide about environmental conditions, rather than to describe what in particular was being measured or how measurements were performed."

Perhaps the most extensive work related to multimetric indicators in the SP literature is that related to the standard water quality ladder or index (McClelland 1974; Vaughan 1986; Mitchell and Carson 1989). This index at its best is linked to specific pollutant levels, which in turn are linked to the presence of aquatic species and suitability for particular recreational uses. An index linked in this way allows the use of objective water quality parameters (e.g., DO concentrations) to characterize uses provided by a given water body. Unfortunately, most studies value the various rungs of the ladder (e.g., suitable for boating, fishing, swimming) without any underlying linkages to water quality parameters.

Dozens of studies have applied variants of the water quality ladder when estimating WTP for water quality improvements (e.g., Johnston et al. 2005a). Although various studies have critiqued the standard version of the ladder (e.g., Johnston et al. 1995; Van Houtven et al. 2007), most of these have offered other multimetric water quality indices as an alternative. Few papers in the SP literature have questioned the central concept of a multimetric water quality index. To the contrary, Bateman et al. (2005) have illustrated how such ladders can increase the understanding and salience of environmental changes among respondents. One of the exceptions is Martin-Ortega and Berbel (2010), who combined a choice experiment and AHP to explore the consistency between the order of uses on a water quality ladder and the relative importance given to these uses within AHP. Their results suggest that the water quality ladder's implied increasing values of uses associated with increasing water quality do not correspond to the relative value of these uses.

These combined results suggest that indicators with emergent properties not only pass the comprehension test but also directly relate to what many people value. Recent empirical work

supports this contention (Johnston et al. 2011, 2013a, 2013b). However, unless the rungs of the ladder are linked to specific biophysical parameters, the usefulness and ecological salience of these indicators are limited. Note that this issue is different from the aggregated versus disaggregated indicators issue discussed under key topic 4 above. Specifically, it suggests that the combined indicator has a fundamental meaning that is distinct from that of the included subindicators. For example, providing lay respondents with information on one or more individual biochemical properties of water will generally be insufficient for them to understand whether the water is suitable for valued uses. Moreover, the effect of some subindicators on valued aspects of water (e.g., the effect of nutrient concentrations on the aesthetic properties of a water body) may depend on other subindicators (e.g., water temperature). This is an important area for future research.

Still more controversial are some less well-studied indicators, for example, multimetric indicators such as indices of biotic integrity noted above (Karr 1981, 1991; Suter 1993; Suter and Glenn 2001). While some research has sought to use indicators of emergent properties within analyses of environmental values (Jakus and Shaw 2003; Johnston et al. 2011, 2013b), use of these indicators by economists and other social scientists is not yet widespread. Hence, while there is increasing consensus that emergent properties can be directly relevant to people (particularly when nonuse values are considered), there is little consensus on the indicators best suited to measure these outcomes or even whether it is possible to measure these outcomes in any meaningful way. Considering this lack of consensus, the treatment of emergent properties within a linking indicators framework remains an important research question.

The issue of multimetric (or cross-category) versus individual indicators manifests differently in the RP literature, with some types of studies relying on more aggregated indicators and other types relying on less aggregated indicators. These differences often relate to the underlying goal of the RP study. For example, recreation demand studies that estimate the total value of access to a recreational site tend to focus on aggregated commodities that can be characterized using a combination of indicators. A beach, for example, is really a collection of characteristics, including water quality, width, temperature, wave action, view, sand characteristics, and species populations. The hedonic property value literature also tends to focus on estimating the property value impacts of more aggregated commodities such as beaches (Bin et al. 2005), forests, cropland, pasture (Irwin 2002), wetlands (Mahan et al. 2000), and open space (Smith et al. 2002). However, other studies in the RP literature evaluate behavior and value changes associated with disaggregated indicators. In the recreation demand literature, for example, these tend to appear in studies estimating values associated with changes or differences

in site characteristics, such as the width of beaches (Parsons et al. 1999; Massey and Parsons 2002), species of fish in a fishery (Freeman 1995), and sand quality (Murray et al. 2001). Other RP studies have estimated values associated with indicators of forest edge in rural areas (Cho et al. 2008), elk habitat (Bastian et al. 2002), and a variety of water quality criteria, including clarity and fecal contamination (McConnell and Tseng 1999), biotic integrity scores and water quality standards violations (US Environmental Protection Agency 2002), and nutrient concentrations (Bockstael and Strand 1987; Bockstael et al. 1989).

Another difference between the RP and SP literature related to the relevance of cross-category versus individual indicators is that SP researchers have the flexibility to use various types of indicators when developing surveys, whether or not historical data are available for these indicators (although current baselines must be available). RP researchers, however, can use only those indicators for which data are available for the relevant time period(s) of analysis. As a result, direct comparisons of cross-category versus individual indicators' performance can be conducted only when data on both types of indicators have been collected for all sites and time periods under analysis. At least partially as a result, there are fewer systematic comparisons of this type in the RP literature.

These limitations notwithstanding, it appears that certain types of studies in the RP literature may be more amenable to specific types of aggregated versus disaggregated indicators. Some types of studies, including site value studies, almost necessarily link estimated value to aggregated commodities (e.g., beaches, fishing sites) rather than more specific indicators. Other types of studies are better suited to the incorporation of individual indicators. However, as mentioned above, there has been minimal testing in the RP literature of whether indicators that aggregate over multiple categories (cross-category indicators, such as ecological health indices) perform better than indicators that focus on specific ecological elements. Currently, researchers make these choices on a case-by-case basis, often based on convenience, data availability, and similar factors. The risk of such ad hoc treatments is that the resulting models may be misspecified (e.g., have measurement error or omitted variables bias), leading to misguided conclusions. Without future research in this area, the pervasiveness of such problems in the literature is unknown. Whether RP models would be improved by more systematic guidance in this area is an important question for future research.

6. *What are the temporal and spatial dimensions of specific ecosystem-beneficiary pairings that matter? Can any generalizations be made about indicator performance along the temporal and spatial dimensions?*

Linking indicators have spatial and temporal dimensions. This applies to those used in both RP and SP studies. Relevant questions in this domain include the following:

- When considering policies with delayed or progressive effects, over what time periods should indicators be presented and aggregated (or averaged)? Under what instances (and how) should indicators communicate outcomes over multiple time periods (versus a long-run equilibrium or average)?
- Over what geospatial areas should indicators be presented and aggregated? Under what instances (and how) is spatial heterogeneity in indicator outcomes relevant? At what degree of resolution should spatial heterogeneity be considered? Under what conditions are discrete (i.e., patchy) versus continuous spatial indicator patterns relevant?

A mature body of literature addresses the temporal stability (e.g., Brouwer 2006; Johnston and Rosenberger 2010; Liebe et al. 2012; Rosenberger and Johnston 2009; Zandersen et al. 2007) and spatial dimensions (e.g., Bateman et al. 2006; Campbell et al. 2009; Johnston and Duke 2009; Johnston and Ramachandran 2014; Loomis 2000; Schaafsma et al. 2012) of economic behavior and values (e.g., how values differ according to the location of individuals relative to ecological changes, with the spatial scale and configuration of these changes, and with extent of the market). These issues are discussed above and are distinct from the relevant spatial and temporal determinants of the linking indicators most closely associated with these behaviors and values. Here the literature is largely (although not totally) silent. However, indirect insights can be drawn from areas of the literature not directly focused on indicators.

Although the answers to these questions often depend on the particular characteristics of the behavior (RP) or potential values or choices (SP) being modeled, the literature has identified some common patterns that can be used to guide indicator development. Moreover, methods such as preliminary focus groups (SP) or interviews with the beneficiary group in question (RP) can often be used to characterize the most relevant spatial and temporal dimensions of various indicators within specific case studies. For example, interviews with farmers in particular areas could be used to evaluate the temporal and spatial aspects of indicators considered relevant for their behavior (e.g., expected rainfall over an entire year or over only specific months). Analyses of farm production (as affected by rainfall) could also be used to address this question. Focus groups or cognitive interviews can be used to evaluate similar issues for responses to SP questionnaires.

Although steps such as these are often part of applied research efforts, their results (e.g., in terms of motivating spatial or temporal aspects of linking indicators) are not often described in

published documents. This lack of description parallels more general limitations in data reporting within valuation studies (Loomis and Rosenberger 2006). Exceptions include such works as Johnston et al. (2002), who detail the use of focus groups to determine the degree of spatial resolution and specificity most useful to SP respondents when considering rural development and conservation plans. Opaluch et al. (1993) provide a similar discussion of the types of local and regional spatial information relevant to respondents when choosing landfill sites, again as revealed by focus groups conducted during SP survey design. Similar approaches can be used to provide input regarding relevant temporal dimensions. For example, Bateman et al. (2005) describe the use of focus groups to determine the time period over which environmental impacts were presented in an SP questionnaire addressing reductions in lake acidity.

To balance concerns of plausibility (policies cannot have immediate impacts) and discounting (policies in the far future are valued very little by respondents), they chose to present policy effects as the "endpoints of a ten year process." Similarly, in his focus group work, one of the authors (Krupnick) found that foreshortening of time for indicators to change is necessary because longer time-period descriptions lead to doubt about whether the program is working and uncertainty about whether experts really understand what is going on with an ecosystem or what is needed to improve it. Another author (Weber) investigated temporal preferences for listed salmon recovery in Oregon (EPA ICR No. 2489.01). Both overly short and overly long timelines were rejected by participants as either unrealistic or too far in the future to make predictions, respectively. Thus, in revising the survey, the SP designers selected timeline options found to be both publicly and scientifically believable.[20] Finally, Van Houtven et al. (2014) report that spatial aspects of algae blooms complicated survey design because of respondents' belief that they could simply alter their activities to avoid areas where blooms were taking place.

The SP literature also provides a variety of indirect evidence regarding temporal dimensions of indicators and outcomes that may be most salient to respondents. Meyer (2013), for example, conducted a direct choice experiment to examine the effects of a delay in

20 One issue in the literature is temporal insensitivity of WTP. Kim and Haab (2009) observed this issue and suggested that instead of using the present value of WTP, where the assumption is that "respondents calculate the willingness to pay for the benefits in each time period of a proposed project and then perform mental gymnastics to discount that stream of values back to the present," one should use total WTP, which "assumes the respondent views the entire stream of benefits as a whole and constructs a value based on the entire stream as perceived at the time of response." However, this comment is about temporal issues in WTP and cost representation (e.g., whether the cost attribute should be portrayed in annual or monthly terms), not about temporal descriptions of linking indicators.

improvement on the benefits of restoration in the Minnesota River Basin. Based on survey responses, he estimated an average discount rate of 13% and calculates that a five-year delay in improvements causes Minnesota residents to lose almost half the benefits of restoration. Viscusi et al. (2008) drew on a large nationally representative sample to estimate discount rates for water quality improvements and found rates of time preference for users to be in the 5 to 10% range and for nonusers in the 17 to 23% range (this again relates to the beneficiary-dependence issue raised in key topic 3). These results indicate that people are sensitive to the timing of improvements and that timing can have a large impact on benefits. Kovacs and Larson's (2008) study of open space in Portland, Oregon, found a discount rate of approximately 30%, indicating that people are highly sensitive to timing with respect to payment for improvements.

These and other works in the literature show that the value of ecological improvements can decline as one moves farther into the future, based on standard (though often high) rates of discount. A related topic is whether studies of similar resources conducted during different time periods show different patterns of value. Findings here are sparse and mixed. Rosenberger and Johnston (2009) provided one of the few existing analyses of temporal trends in nonmarket values across different types of environmental outcomes, showing that these trends differ in both magnitude and sign across different resource types, confounding the search for universal patterns. These and other works on temporal dimensions of nonmarket values offer clear evidence that temporal issues are relevant, but they do not indicate to what extent these issues are directly relevant to the choice of linking indicators. Additional research is required, therefore, to identify the extent to which temporal dimensions inform the choice of variables to be monitored or modeled by natural scientists.

The literature addressing spatial aspects of nonmarket values similarly provides substantial indirect evidence, but it affords little direct insight into the most relevant spatial dimensions of linking indicators. The literature demonstrates clearly that the value of otherwise identical ecological outcomes can vary across different geographic sites and over spatial distance (Bateman et al. 2006, 2011b; Campbell et al. 2009; Johnston and Rosenberger 2010; Schaafsma et al. 2012; Jorgensen et al. 2013; Johnston and Ramachandran 2014). These differences are often large, leading to large errors if value estimates from one area are used to approximate those in other areas (Rosenberger and Stanley 2006) or if the impact of spatial factors is ignored (Bateman et al. 2006; Johnston and Ramachandran 2014). Adjusting for these differences and minimizing attendant errors are among the primary goals of the benefit transfer literature in applied economics (Johnston and Rosenberger 2010). In extreme cases, ecological outcomes that are socially relevant (or have nonzero value) in some areas may be irrelevant in others. More

commonly, the most relevant or highest-valued ecological outcomes vary across regions. For example, Campbell et al. (2009) demonstrated that values for otherwise identical rural landscape changes vary spatially across the Republic of Ireland. Johnston and Ramachandran (2014) illustrated similar findings for otherwise identical biophysical outcomes of fish passage restoration in Rhode Island. Because the social relevance of otherwise identical biophysical outcomes varies across people and space, it follows that the most important linking indicators will also differ.

There is also evidence that the most relevant environmental outcomes vary according to the geographic scale over which these outcomes are evaluated. For example, Johnston and Duke (2009) found that relative values for different types of farmland attributes (e.g., land and crop type) vary depending on whether one evaluates values for preservation at the community or state scale. Hein et al. (2006) provided a similar conceptual argument, describing a framework based on a literature review for ecosystem valuation suggesting that ecosystem services accrue to beneficiaries at different scales, and therefore different goods will be relevant to people depending on the scale. For example, a small-scale fishery may not be salient to someone at a global scale, but it has high salience to people at the local scale, especially to fishermen. This point suggests that spatial scale is important for defining linking indicators (i.e., attributes are not necessarily transferable between scales). What works at one scale may not work at another scale.

The literature has also found that microlevel spatial aspects of environmental outcomes presented to respondents are relevant, apart from effects related to respondents' location (e.g., distance decay in values). Johnston et al. (2002), for example, found that spatial elements of rural development and conservation plans shown on SP survey maps were directly relevant to responses, even when these spatial aspects were not highlighted in text. Van Houtven et al. (2014) also found that spatial aspects of algae blooms were directly relevant to survey design.

The RP literature is difficult to evaluate with regard to the implications of spatial factors for the most relevant linking indicators. As discussed by Bateman et al. (2002), Bockstael (1995), and Bockstael and McConnell (2010), much of the focus of RP analyses, as in other areas of environmental and resource economics, is inherently spatial. The hedonic property value literature, for example, is fundamentally spatial, with hundreds of articles over the last four decades estimating the effects of spatially proximate and differentiated environmental changes on property values (Geoghegan et al. 1997; Paterson and Boyle 2002; Mueller and Loomis 2008). Similarly, recreational site choice is fundamentally influenced by spatial factors, including site distance and spatial relationships between alternative sites (Parsons et al. 1999). Yet, as in the SP literature, analyses in this literature have not generally focused on the

implications of spatial factors for the relevance of different indicators. For example, many works in the hedonic literature have estimated distances over which the effect of environmental changes on property values can be discerned. These typically vary over different types of effects and different areas. To the knowledge of the authors, however, these findings have not been synthesized into general findings related to the general relevance of different types of indicators as influenced by spatial considerations. Significant insight could be provided through more systematic analysis of this literature (e.g., meta-analysis) with a specific focus on this issue.

In summary, to adequately characterize final outcomes and linking indicators, it is necessary to understand how they vary across different sites, populations, and temporal scales. That is, identification of linking indicators requires that ecosystems be understood "in their particular spatially and temporally defined context" (Bateman et al. 2011a). This again is a research question for which a large foundational literature exists, but for which significant research and consensus are nonetheless required, including both primary research and synthesis of prior results. Among the primary challenges of this research is to identify general guidelines for ecological monitoring, modeling, and mapping, while allowing for the fact that socially relevant ecological outcomes vary across regions, populations, and temporal contexts. This occurs because both underlying ecological processes and linkages between humans and ecosystems vary geographically and over time.

7. *Does the existence (or nonuse) value context present any specific complications in indicator design relative to the use value context?*

Nonuse values are critical elements of ecosystem values. They may be broadly defined as values that cannot be measured using data on observed behavior—they are values that do not require observable "use" of a resource or ecosystem (Freeman et al. 2014). Nonuse values have been described as existence values (i.e., they are the WTP of people for a given change in a resource that is unrelated to their use of that resource). For people who do not and never will use the resource, this concept is clear. For users, it is still possible to hold values for that resource over and above any use values they have. Studies show that although nonuse or nonuser values are often relatively low on a per-person basis, these values can potentially be held by a much greater number of people than is typical for most types of use values, such that these values are large in the aggregate (Freeman et al. 2014). For example, the meta-analysis of Johnston et al. (2003) addressing relationships between use and nonuse values for water quality improvements shows that nonuse values can remain positive even when use values are negligible. Similarly, Hanley et al.'s (2003) study of WTP for river flow improvements shows that "more rapid distance decay exists for use values than for non-use values," implying that nonuse values

remain positive at a greater distance from the affected rivers (and are thus positive for a larger population than use values). Hence, the aggregate magnitude of nonuse values can outweigh that of use values. This question asks whether different types of linking indicators are relevant to the estimation of these values or to total WTP estimates dominated by these values. By definition, nonuse values are measurable only using SP studies, so this question is relevant only to this type of analysis.

The literature supports the idea that cognitive and informational burdens are often higher for nonuse values than for use values. This difference may be due to the comparative lack of experience with environmental goods and services that generate (primarily) nonuse values compared with those that generate (primarily) use values. However, this difference in itself does not imply that the most relevant linking indicators necessarily differ. Identification of the most relevant effects of ecological change on humans is often relatively straightforward for the case of use values, defined as values related to observable behaviors. In such cases, one can directly observe activities such as market purchases and uses of nonmarket natural resources, for example. Analysis of these observable behaviors can often provide significant insight into the role and value of ecological outcomes. The subfield of RP valuation in economics is devoted to the study of such behaviors and values (Bockstael and McConnell 2010; Freeman et al. 2014). Through analyses such as these, it is often possible to determine empirically whether and how different types of ecological changes influence human behavior. Comparisons of studies across different policy contexts can provide insight into ecological measures that are most closely associated with human behavior and use value. These issues are discussed under key topics 1 through 6 above.

Such evaluations are often considerably more difficult for the case of nonuse values. Although certain general expectations exist regarding the types of ecological outcomes that might be associated with large or small nonuse values, there are few theoretically necessary rules (Carson et al. 1999; Freeman et al. 2014). According to economic theory, a person can hold a nonuse value for almost anything. For example, it is theoretically possible that certain people might simply value a reduction in N concentrations in rivers for no other reason than pure existence or bequest value. Hence indicators that would be very unlikely to serve as suitable linking indicators in a use value context could potentially be linking indicators in a nonuse value context. As in the case of RP methods above, empirical evaluations are necessary to determine whether and how different types of ecological changes influence nonuse values.

Unlike the evaluation of use values, however, observable human behaviors do not provide any special insight into nonuse values. Formal, quantitative insight into these values and

their determinants can be gained only through carefully crafted SP methods (Carson et al. 1999;, Bateman et al. 2002; Freeman et al. 2014), with survey design input from qualitative research methods such as focus groups and cognitive interviews (Desvousges and Smith 1988; Johnston et al. 1995; Kaplowitz et al. 2004; Powe 2007). As with many of the questions addressed above, the literature provides only indirect evidence. For example, Johnston et al. (2005b) have shown that WTP for changes in wetland attributes differ between users and nonusers for some (including the relative abundance of wetland-dependent birds and fish) but not all types of wetland attributes (e.g., the value of shellfish improvements is unaffected).

The fact that nonuse values can be held for almost any type of ecological change, combined with the great variation in the (often ambiguous) use of ecological indicators to quantify "relevant" ecological outcomes in past economics research (Schultz et al. 2012; Zhao et al. 2013), leads to ambiguity concerning the linking indicators that might be most closely related to nonuse values in different policy contexts. Indeed, Zhao et al. (2013) demonstrated that some types of ecological indicators can be considered formally "outcome-equivalent" for purposes of economic valuation; they define outcome-equivalent indicators as different indicators that communicate the status of identical underlying ecological outcomes, as perceived by respondents. This finding is not necessarily general, however; in many other instances, different indicators communicate different valued outcomes—or communicate the same outcome with different degrees of precision, immediacy, and comprehensibility (Johnston et al. 2012; Schultz et al. 2012). Hence before a linking indicators framework can be applied in practice, additional work (including research synthesis or meta-analysis) will be required to help identify linking indicators most closely associated with ecological outcomes that provide substantial nonuse value. As noted above, these may differ considerably across areas, beneficiary groups, and time.

In conclusion, we find that the identification of linking indicators is a more open-ended process in a nonuse value context than in a use value context, with fewer clear guidelines regarding appropriate indicators. For example, indicators that would almost certainly *not* be linking indicators in most use value contexts (e.g., nutrient concentrations in water) could potentially serve as linking indicators in a nonuse value context. The challenge identifying linking indicators for nonuse values is further compounded by the lack of generalizable, direct evidence in the literature and by the fact that RP studies provide no insight in this area. In summary, this is an area in which the least is known about the characteristics of linking indicators and in which research is most needed. This lack of evidence notwithstanding, the fundamental properties of linking indicators remain similar whether in a use or nonuse value context.

8. *Can any generalizations be made about the unit of account to express a linking indicator (e.g., acres, percentage change, absolute numbers, qualitative indicators [high, medium, low], icons, pictures)?*

Indicators generally have three dimensions beyond their category (e.g., fish populations): space, time, and unit of account. This question addresses the last of these. One of the only papers to address this question from a general perspective is Schultz et al. (2012), which proposes four standards for indicators used in SP analysis, two of which are directly relevant to indicator units of account. First, indicators should be empirically measurable. This is defined as indicators that "have a clearly stated relationship to ecological data or model results [and] consist of measures that are, at a minimum, empirically quantifiable" (304). This standard excludes categorical units of account such as "high, medium, and low" or graphic icons, unless these categories are clearly linked to underlying measurable indicators. The challenge here is not the use of categorical indicators, but rather the lack of grounding of these categories in clearly defined ecological measurements, so that the categories have no precise interpretation. Second, indicators should be interpretable. This implies that "different possible values for indicators should be understood similarly by respondents and scientists" (304). At the same time, Bateman et al. (2005), Johnston et al. (1995), Opaluch et al. (1993), and many others argue that the communication of quantitative information in SP surveys is often enhanced through graphic elements such as icons, ladders, and images. These discussions highlight two aspects of "units of account" that are relevant for the development of linking indicators. First, units of account must be such that indicators have unambiguous interpretations. Second, they should enable indicators to be understood similarly by laypersons and experts. This understanding can sometimes be enhanced by presenting indicators in multiple ways (e.g., using cardinal measurements and graphic illustrations).

The SP literature also provides significant case study evidence regarding the relevance of different units of account. For example, one of the biggest early controversies in SP studies was associated with research related to the Exxon Valdez oil spill, in which certain units of account (e.g., the number of dead birds) was shown to lead to responses that were insensitive to scope. Subsequent studies indicate that this unit of account was poorly chosen (Hanemann 1994; Carson 2012). Various authors, including Hanemann (1994) and Johnston et al. (1995), have suggested that a likely explanation of this result is that people placed both numbers in the same category: a very small percentage of the total number of birds. Recent analyses have placed more emphasis on using percentages or other indicators of relative change (e.g., relative to ecological reference conditions), either instead of or in addition to indicators communicating cardinal changes

(Johnston et al. 2012). For example, research by Luisetti et al. (2011) indicates that respondents may perceive attribute levels (here, the distance to a restored wetland) in relative terms as opposed to absolute terms. This complicates valuation, as it implies that values might be sensitive to the selected reference conditions.

Indicators can also involve critical thresholds. For example, Adamowicz et al. (1998) found that where thresholds matter, there will be scope insensitivity to numerical or percentage changes that improve a population beyond the threshold. Specifically, the information stated that the current caribou population was 400, but biologists suggested that a minimum population of 600 is required for a viable population with a small risk of extinction. The authors noticed that the marginal utility of caribou declines dramatically after a population of 600 has been reached in a specific area. This finding, of course, is relevant to the specific area and survey sample studied—other types of thresholds (or a lack of thresholds) might be relevant to other contexts or beneficiary groups. Similar threshold issues are discussed by Heberlein et al. (2005), Bateman et al. (2005), and Weber and Stewart (2008). For example, the Bateman et al. analysis suggests that UK households are willing to pay a significant amount to protect a minimum threshold number of lakes, but after that threshold is reached, the WTP to protect additional lakes is modest.

Within the RP literature, the unit of account issue is less central. Given the nature of RP analysis, all indicators used must be quantitative and measurable, to be included within parametric or nonparametric analysis. As discussed above, the most appropriate of the available units (or resolution) of measurement is generally determined via comparative performance in statistical models. Moreover, also as discussed above, RP analysts are generally limited to the indicators already available for their study sites and hence have less freedom to choose the units of account best suited to modeling behavior in particular circumstances.

Summary

From the above literature review, it is clear that various areas of theoretical and empirical literature provide insight into the characteristics of linking indicators across various dimensions, as reflected in our eight key topics. However, while some general guidance can be derived from this literature, the evidence related to most of the questions they ask can best be characterized as scattered and indirect. Little attention has been focused on the specific characteristics of indicators best suited for communicating ecological or environmental changes that are most directly relevant to the public and best suited for coupled natural and social science evaluations. This lack of attention—and the tendency of monitoring and other programs to measure nonlinking indicators—is among the primary motivations for this paper. Because much of the

relevant literature does not focus primarily on indicator definition, the insight from this literature is suggestive but not conclusive.

Among the general conclusions that can be drawn from the literature is that the properties of indicators can—but need not always—have important implications for the outcomes of behavioral and valuation research. However, while empirical evaluations show that results can vary according to indicator properties, these evaluations do not frequently indicate which set of indicators is "best." Moreover, while certain conclusions have strong support in the literature (e.g., that spatial and temporal factors are relevant), other questions have little supporting evidence (e.g., what indicators are best suited to evaluating nonuse values). Still other areas of uncertainty reflect broader areas of controversy in the ecological literature (e.g., the usefulness and interpretability of multimetric indicators). The answers to other questions (e.g., relationships between indicator aggregation and performance in social science models; the degree to which indicators should be targeted to distinct beneficiary groups) appear to be largely context dependent, although evidence suggests that greater targeting is appropriate for users with greater experience levels.

Considering the current state of the literature, at least two types of research are needed to address knowledge gaps related to linking indicators. First, additional synthesis is required to draw more systematic and general conclusions from the literature both in- and outside of valuation regarding the actual and potential performance of different types of indicators in different RP and SP contexts. Second, additional (new) qualitative and quantitative empirical work is required to evaluate and compare the performance of different types of indicators in different situations. While some research has begun to focus explicitly on the use and definition of indicators within valuation and behavioral modeling, this is still an immature and developing area of work. As noted by Johnston et al. (2012), the ubiquity of "ad hoc characterizations of ecological outcomes" in the valuation literature suggests the critical need for work and guidance in this area.

6. Conclusions and Recommendations

Collaboration between natural and social scientists is necessary to depict nature's contributions to social well-being and evaluate the benefits of environmental investments, management, and policy. A key to collaboration is the identification and measurement of *linking indicators*, biophysical indicators that facilitate social evaluation and can be directly linked to human welfare.

As noted earlier, the selection and measurement of biophysical indicators have traditionally been driven by natural scientists with relatively little input from the social sciences and as a function (understandably) of measurement cost considerations, regulatory mandates, and ecological theory. For their part, social scientists have traditionally underappreciated the importance of biophysical indicators' definition to their own analyses' accuracy and relevance. But government, nongovernmental organization, and business-sector demand for coordinated analysis—and its corollary intellectual underpinnings in the ecosystem services movement—has revealed the need to develop better linking measures. That is, given the scarce resources available to measure, model, and monitor ecological and environmental conditions, it is necessary to focus these efforts on indicators that have demonstrable biophysical AND social relevance.

The broader deployment of linking indicators in monitoring programs, policy analysis, public reporting, and environmental accounting requires natural and social scientists to share a common understanding of their motivation, features, and practical deployment. This report is geared toward and carries implications for both. The paper does the following:

- develops a set of principles to guide the further identification of linking indicators;
- describes linking indicators' role in various forms of analysis;
- compares their features with those of more commonly collected ecological measures; and
- reviews empirical evidence pertinent to identifying and defining linking indicators.

For any team of natural resource or environmental policy evaluators, we recommend, at a minimum, conceptual development of an ecological production framework that describes causal linkages among biophysical outcomes in an ecological system and ultimately leads to impacts on different beneficiary groups. Ecological production frameworks have several virtues. First, ecological production frameworks' identification of causal relationships among ecological outcomes can facilitate the identification of diverse beneficiaries affected by changes in the ecosystem. Our review of the empirical literature highlights the significant degree to which natural resources' value can depend on an outcome's location, type of use or enjoyment, timing, and other beneficiary-specific factors. Ecosystem beneficiaries are diverse and draw value from nature in diverse ways, from consumptive natural resource uses to recreation, aesthetic enjoyment, and ethical and stewardship motivations. This range generates a corresponding diversity in linking biophysical measures. Accordingly, an initial broad recommendation is that analysts identify linking indicators in reference to what may be heterogeneous (e.g., demographically, geographically) sets of stakeholders affected by natural resource conditions.

Production frameworks such as these also permit ecological outcomes (and their indicators) to be differentiated based on the degree to which they matter directly versus indirectly to beneficiary welfare. Such differentiation is important because, as a baseline hypothesis, indicators that are more proximate to social welfare will (1) be more meaningful and understandable to lay audiences; and (2) lead to more accurate and interpretable monetary valuations of ecological outcomes. Finally, production frameworks can help identify and organize the full set of models, expertise, and data needed to relate intermediate (or "distal") ecological outcomes to more proximal, linking indicators and to resource management options, stressors, and conservation actions. In summary, production frameworks can help elucidate the causal chain through which ecological changes influence people, and through this understanding, they can promote the measurement of indicators that are more relevant (or proximate) to human effects.

Our review of the literature suggests that the desirability of proximal outcomes as linking indicators is empirically understudied and should therefore be treated as a strong theoretical hypothesis rather than as an unequivocal, generalizable fact. We strongly advocate more explicit empirical testing of the hypothesis. Assuming our hypothesis is borne out, our second broad recommendation is that analysts identify via ecological production frameworks and measure biophysical outcomes as close as possible to beneficiary welfare, given real-world measurement and modeling constraints.

It deserves emphasis that the relative desirability of alternative indicators (in terms of their ability to communicate or contribute to more accurate social evaluations) can be evaluated empirically. Our review of the empirical literature describes several ways in which this can be done. But despite that capability, the review also reveals that relatively little attention has been given to the subject, even by empirically sophisticated practitioners. Open questions identified by the review, such as the most appropriate units of account, most appropriate nonuse indicators, and approaches to aggregation, would benefit greatly from more deliberate empirical examination. Such studies would not only help improve the accuracy and salience of ecosystem services assessments but also lead to greater consensus among practitioners on preferred indicators. Greater consensus would advance the standardization of indicator protocols—a desirable goal because of the need to compare, aggregate, integrate, and transfer monitoring and evaluation results across the nation's ecological and social landscapes. It would also help standardize valuation approaches around more directly relevant indicators—a need emphasized by Schultz et al. (2012) and others.

For those seeking practical guidance on what to measure, we make the observation that the answer is unavoidably a function of the specific issue being evaluated and the financial and technical resources that can be brought to bear. However, for any application seeking coordinated natural and social analysis, we recommend attention by both natural and social science analysts to the relevant ecological production system, its corollary identification of relevant beneficiaries, and emphasis on outcomes as proximate as practicable to beneficiary welfare.

In our view, the next step in linking indicator development involves selecting concrete applications on which a set of related research activities could be based. Starting with a case-specific ecological production system codeveloped by natural scientists, social scientists, and beneficiaries, the cases could be used as a framework for analysis of (1) current and potential linking indicator monitoring capacity and costs; (2) beneficiary-specific indicator development; and (3) empirical evaluation of alternative indicators and indicator formats. The outcome of such efforts could be a significant advancement toward measurement and monitoring protocols applicable beyond the specific cases.

References

Adamowicz, W., P. Boxall, M. Williams, and J. Louviere. 1998. Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. American Journal of Agricultural Economics **80**:64–75.

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams. 1997. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. Journal of Environmental Economics and Management **32**:65–84.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. 1993. Report of the NOAA panel on contingent valuation. Federal Register 58: 4601-4614

Baker, J. 1991. NAPAP Report 13: biological effects of changes in surface water acid-base chemistry. Page 381 *in* P. M. Irving, editor. Acidic Deposition: State of Science and Technology. National Acid Precipitation Assessment Program, Washington, DC.

Baker, J., and S. Christensen. 1991. Effects of acidification on biological communities in aquatic ecosystems. Pages 83–106 *in* D. Charles, editor. Acidic Deposition and Aquatic Ecosystems. Springer, New York.

Banzhaf, H. S., and J. Boyd. 2012. The architecture and measurement of an ecosystem services index. Sustainability **4**:430–461.

Banzhaf, H. S., D. Burtraw, D. Evans, and A. Krupnick. 2006. Valuation of natural resource improvements in the Adirondacks. Land Economics **82**:445–464.

Barbier, E. B., S. Baumgartner, K. Chopra, C. Costello, A. Duraiappah, R. Hassan, A. P. Kinzig, M. Lehman, U. Pascual, S. Polasky, and C. Perrings. 2009. The valuation of ecosystem services. Page 368 *in* S. Naeem, D. E. Bunker, A. Hector, M. Loreau, and C. Perrings, editors. Biodiversity, Ecosystem Functioning, & Human Wellbeing. Oxford University Press, Oxford, UK.

Bastian, C. T., D. M. McLeod, M. J. Germino, W. A. Reiners, and B. J. Blasko. 2002. Environmental amenities and agricultural land values: a hedonic model using geographic information systems data. Ecological Economics **40**:337–349.

Bateman, I. J., R. Brouwer, S. Ferrini, M. Schaafsma, D. N. Barton, A. Dubgaard, B. Hasler, S. Hime, I. Liekens, and S. Navrud. 2011b. Making benefit transfers work: deriving and testing principles for value transfers for similar and dissimilar sites using a case study of

the non-market benefits of water quality improvements across Europe. Environmental and Resource Economics **50**:365–387.

Bateman, I. J., R. T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, and E. Özdemiroglu. 2002. Economic valuation with stated preference techniques: a manual. Edward Elgar, Northampton, MA.

Bateman, I., P. Cooper, S. Georgiou, S. Navrud, G. Poe, R. Ready, P. Riera, M. Ryan, and C. Vossler. 2005. Economic valuation of policies for managing acidity in remote mountain lakes: examining validity through scope sensitivity testing. Aquatic Sciences—Research across Boundaries **67**:274–291.

Bateman, I. J., B. H. Day, S. Georgiou, and I. Lake. 2006. The aggregation of environmental benefit values: welfare measures, distance decay and total WTP. Ecological Economics **60**:450–460.

Bateman, I., G. Mace, C. Fezzi, G. Atkinson, and K. Turner. 2011a. Economic analysis for ecosystem service assessments. Environmental and Resource Economics **48**:177–218.

Bateman, I. J., A. Munro, and G. L. Poe. 2008. Decoy effects in choice experiments and contingent valuation: asymmetric dominance. Land Economics **84**:115–127.

Bennett, J., and V. Adamowicz. 2001. Some fundamentals of environmental choice modelling. Pages 37–69 *in* J. Bennett and R. Blamey, editors. The Choice Modelling Approach to Environmental Valuation. Edward Elgar, Cheltenham UK.

Bernard, H. R. 2011. Research methods in anthropology: Qualitative and quantitative approaches. Rowman Altamira, Lanham, MD.

Bin, O., C. E. Landry, C. L. Ellis, and H. Vogelsong. 2005. Some consumer surplus estimates for North Carolina beaches. Marine Resource Economics **20**:145.

Bishop, R. C. 2003. Where to from here? Pages 537–566 *in* P. A. Champ, J. Boyle, and T. C. Brown, editors. A Primer on Nonmarket Valuation. Kluwer, Dordrecht, Netherlands.

Bishop, R. C., K. J. Boyle, and M. P. Welsh. 1987. Toward total economic valuation of great lakes fishery resources. Transactions of the American Fisheries Society **116**:339–345.

Blamey, R. K., J. W. Bennett, J. J. Louviere, M. D. Morrison, and J. C. Rolfe. 2002. Attribute causality in environmental choice modelling. Environmental and Resource Economics **23**:167–186.

Bockstael, N., R. Costanza, I. Strand, W. Boynton, K. Bell, and L. Wainger. 1995. Ecological economic modeling and valuation of ecosystems. Ecological Economics **14**:143–159.

Bockstael, N. E., and K. E. McConnell. 2007. Environmental and resource valuation with revealed preferences: a theoretical guide to empirical models. Springer, Dordrecht, Netherlands.

Bockstael, N. E., K. E. McConnell, and I. E. Strand. 1989. Measuring the benefits of improvements in water quality: the Chesapeake Bay. Marine Resource Economics **6**:1–18.

Bockstael, N. E., and I. E. Strand. 1987. The effect of common sources of regression error on benefit estimates. Land Economics **63**:11–20.

Boyd, J. 2007. Nonmarket benefits of nature: what should be counted in green GDP? Ecological Economics **61**:716–723.

Boyd, J., and S. Banzhaf. 2007. What are ecosystem services? The need for standardized environmental accounting units. Ecological Economics **63**:616–626.

Boyd, J. W., and A. J. Krupnick. 2013. Using ecological production theory to define and select environmental commodities for nonmarket valuation. Agricultural and Resource Economics Review **42**:1–32.

Brouwer, R. 2006. Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event. Ecological Economics **60**:399–406.

Brouwer R., I. H. Langford, I. Bateman, and K. Turner. 1999. A meta-analysis of wetland contingent valuation studies. Regional Environmental Change **1**(1): 47–57.

Brouwer, R., J. Martin-Ortega, and J. Berbel. 2010. Spatial preference heterogeneity: a choice experiment. Land Economics **86**:552–568.

Brown, N. R., and R. S. Siegler. 1993. Metrics and mappings: A framework for understanding real-world quantitative estimation. Psychological Review **100**:511.

Brown, T. C., J. C. Bergstrom, and J. Loomis. 2007. Defining, valuing, and providing ecosystem goods and services. Natural Resources Journal **47**:329–376.

Burgan, B., & Engle, V. 2006. National Estuary Program Coastal Condition Report: US Environmental Protection Agency 842. B-06/001.

Burtraw, D., A. Krupnick, E. Mansur, D. Austin, and D. Farrell. 1997. The costs and benefits of reducing acid rain. Discussion paper. Resources for the Future, Washington, DC.

Cai, B., T. A. Cameron, and G. R. Gerdes. 2011. Distal order effects in stated preference surveys. Ecological Economics **70**(6):1101–1108.

Cairns, J., P. McCormick, and B. Niederlehner. 1993. A proposed framework for developing indicators of ecosystem health. Hydrobiologia **263**:1–44.

Cameron, T. A., and J. Englin. 1997. Respondent experience and contingent valuation of environmental goods. Journal of Environmental Economics and Management **33**:296–313.

Campbell, D., W. G. Hutchinson, and R. Scarpa. 2009. Using choice experiments to explore the spatial distribution of willingness to pay for rural landscape improvements. Environment and Planning A **41**:97–111.

Carson, R. T. 2012. Contingent valuation: a practical alternative when prices aren't available. Journal of Economic Perspectives **26**:27–42.

Carson, R. T., N. E. Flores, and R. C. Mitchell. 1999. The theory and measurement of passive-use value. Pages 97-130 *in* I. J. Bateman and K. G. Willis, editors. Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries. Oxford University Press, Oxford, UK.

Carson, R. T., and R. C. Mitchell. 1995. Sequencing and nesting in contingent valuation surveys. Journal of Environmental Economics and Management **28**:155–173.

Champ, P. A., K. J. Boyle, and T. C. Brown, eds. Series editor I. J. Bateman. 2003. A Primer on Nonmarket Valuation. Kluwer Academic Publishers, Dordrecht, Netherlands.

Chang, H., P. Thiers, N. R. Netusil, J. A. Yeakley, G. Rollwagen-Bollens, S. M. Bollens, and S. Singh. 2014. Relationships between environmental governance and water quality in a growing metropolitan area of the Pacific Northwest, USA. Hydrology and Earth System Sciences **18**:1383–1395.

Chestnut, L. G., and D. M. Mills. 2005. A fresh look at the benefits and costs of the US acid rain program. Journal of Environmental Management **77**:252.

Cho, S.-H., N. C. Poudyal, and R. K. Roberts. 2008. Spatial analysis of the amenity value of green open space. Ecological Economics **66**:403–416.

Christie, M., N. Hanley, J. Warren, K. Murphy, R. Wright, and T. Hyde. 2006. Valuing the diversity of biodiversity. Ecological Economics **58**:304–317.

Clayton, S., and G. Myers. 2009. Conservation Psychology: Understanding and Promoting Human Care for Nature. John Wiley & Sons Ltd, Chichester, West Sussex, UK.

Council on Environmental Quality. 1997. The 1997 Report of the Council on Environmental Quality. Washington, DC.

Creswell, J. W. 2013. Qualitative Inquiry and Research Design: Choosing among Five Approaches. 2nd ed. Sage Publications, Thousand Oaks, CA.

Creswell, J. W. 2014. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 4th Edition. Sage, Los Angeles, CA.

Czajkowski, M., N. Hanley, and J. LaRiviere. 2015. The effects of experience on preferences: theory and empirics for environmental public goods. American Journal of Agricultural Economics **97**:333–351.

Daily, G. C., S. Polasky, J. Goldstein, P. M. Kareiva, H. A. Mooney, L. Pejchar, T. H. Ricketts, J. Salzman, and R. Shallenberger. 2009. Ecosystem services in decision making: time to deliver. Frontiers in Ecology and the Environment **7**:21–28.

Dale, V. H., and S. C. Beyeler. 2001. Challenges in the development and use of ecological indicators. Ecological Indicators **1**:3–10.

Daw, T., K. Brown, S. Rosendo, and R. Pomeroy. 2011. Applying the ecosystem services concept to poverty alleviation: the need to disaggregate human well-being. Environmental Conservation **38**:370–379.

deHayr, R., and I. Gordon. 2004. Irrigation Water Quality: Salinity and Soil Structure Stability. Queensland Government, Natural Resources, Mines and Energy.

Dernburg, T., & McDougall, D. 1972. Macroeconomics: The Measurement. Analysis, and Control of Aggregate Economic Activity. Mc-Graw Hill, New York.

Desvousges, W. H., and V. K. Smith. 1988. Focus groups and risk communication: the “science” of listening to data. Risk Analysis **8**:479–484.

Dillman, D. A., J. D. Smyth, and L. M. Christian. 2014. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. John Wiley and Sons, New York, NY.

Duke, J. M., and R. Aull-Hyde. 2002. Identifying public preferences for land preservation using the analytic hierarchy process. Ecological Economics **42**:131–145.

Egan, K. J., J. A. Herriges, C. L. Kling, and J. A. Downing. 2009. Valuing water quality as a function of water quality measures. American Journal of Agricultural Economics **91**:106–123.

Englin, J. E., T. A. Cameron, R. E. Mendelsohn, G. A. Parsons, and S. A. Shankle. 1991. Valuation of Damages to Recreational Trout Fishing in the Upper Northeast due to Acidic Deposition. Pacific Northwest National Laboratory, Richland, WA.

Federal Register. 2013. EPA-HQ-OA-2012-0033. 78 FR 9045.

Fern, E. F. 2001. Advanced Focus Group Research. Sage, Thousand Oaks, CA.

Finn, A., and J. J. Louviere. 1992. Determining the appropriate response to evidence of public concern: the case of food safety. Journal of Public Policy and Marketing **11**(2): 12–25.

Fisher, B., and R. K. Turner. 2008. Ecosystem services: classification for valuation. Biological Conservation **141**:1167–1169.

Fisher, B., K. Turner, M. Zylstra, R. Brouwer, R. D. Groot, S. Farber, P. Ferraro, R. Green, D. Hadley, J. Harlow, P. Jefferiss, C. Kirkby, P. Morling, S. Mowatt, R. Naidoo, J. Paavola, B. Strassburg, D. Yu, and A. Balmford. 2008. Ecosystem services and economic theory: integration for policy-relevant research. Ecological Applications **18**:2050–2067.

Fisher, W., L. Jackson, G. Suter, and P. Bertram. 2001. Indicators for human and ecological risk assessment: US Environmental Protection Agency perspective. Human and Ecological Risk Assessment **7**:961–970.

Flynn, T. N., J. J. Louviere, T. J. Peters, and J. Coast. 2007. Best-worst scaling: what it can do for health care research and how to do it. Journal of Health Economics 26(1): 171–189.

Frank, S., C. Fürst, L. Koschke, A. Witt, and F. Makeschin. 2013. Assessment of landscape aesthetics—validation of a landscape metrics-based assessment by visual estimation of the scenic beauty. Ecological Indicators **32**:222–231.

Freeman III, A. M. 1995. The benefits of water quality improvements for marine recreation: a review of the empirical evidence. Marine Resource Economics **10**:385–406.

Freeman, A. M., J. A. Herriges, and C. L. Kling. 2014. The Measurement of Environmental and Resource Values: Theory and Methods. 3rd ed. Taylor and Francis, Abingdon, Oxon, UK.

Geoghegan, J., L. A. Wainger, and N. E. Bockstael. 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. Ecological Economics **23**:251–264.

Giraud, K. L., J. B. Loomis, and R. L. Johnson. 1999. Internal and external scope in willingness-to-pay estimates for threatened and endangered wildlife. Journal of Environmental Management **56**:221–229.

Green, R. H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. Wiley-Interscience, New York, NY.

H. John Heinz III Center for Science, Economics, and the Environment. 2008. The State of the Nation's Ecosystems 2008: Measuring the Land, Waters, and Living Resources of the United States. Island Press, Washington, DC.

Hanemann, W. M. 1994. Valuing the environment through contingent valuation. Journal of Economic Perspectives **8**:19–43.

Hanemann, W. M. 2006. The economic conception of water. Pages 61–92 *in* P. P. Rogers, M. R. Llamas, and L. Martínez-Cortina, editors. Water Crisis: Myth or Reality? Marcelino Botin Water Forum 2004. Taylor and Francis, London.

Hanley, N., and A. Munro. 1992. The effects of information in contingent markets for environmental goods: A survey and some new evidence. Queen's Economics Department Working Paper No. 848. Kingston, Ontario, Canada.

Hanley, N., F. Schläpfer, and J. Spurgeon. 2003. Aggregating the benefits of environmental improvements: distance-decay functions for use and non-use values. Journal of Environmental Management **68**:297–304.

Heal, G. M., E. B. Barbier, K. J. Boyle, A. P. Covich, S. P. Gloss, C. H. Hershner, J. P. Hoehn, S. Polasky, K. Segerson, and K. Shrader-Frechette. 2004. Valuing Ecosystem Services: Toward Better Environmental Decision-Making. National Academy of Science, Washington, DC.

Heberlein, T. A., M. A. Wilson, R. C. Bishop, and N. C. Schaeffer. 2005. Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management **50**:1–22.

Hein, L., K. van Koppen, R. S. de Groot, and E. C. van Ierland. 2006. Spatial scales, stakeholders and the valuation of ecosystem services. Ecological Economics **57**:209–228.

Heiskary, S. A., and W. W. Walker Jr. 1988. Developing phosphorus criteria for Minnesota lakes. Lake and Reservoir Management **4**:1–9.

Hoehn, J. P., F. Lupi, and M. D. Kaplowitz. 2003. Untying a Lancastrian bundle: valuing ecosystems and ecosystem services for wetland mitigation. Journal of Environmental Management **68**:263–272.

Holme, N. A., and A. D. McIntyre. 1971. IBP Handbook No. 16: Methods for the Study of Marine Benthos. Blackwell Scientific Publications, Oxford.

Hovardas, T., and K. Korfiatis. 2011. Towards a critical re-appraisal of ecology education: scheduling and educational intervention to revisit the "balance of nature" metaphor. Science and Education 20:1039–1053.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, S. S. Intelmann, S. C. Corbett, M. C. Arbogast, and R. C. Hjort. 2002. Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. North American Journal of Fisheries Management **22**:1229–1240.

Hughes, R. M., P. R. Kaufmann, and M. H. Weber. 2011. National and regional comparisons between Strahler order and stream size. Journal of the North American Benthological Society **30**:103–121.

Hughes, R. M., and D. V. Peck. 2008. Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics, and reality. Journal of the North American Benthological Society **27**:837–859.

Hunt, L. M. 2005. Recreational fishing site choice models: insights and future opportunities. Human Dimensions of Wildlife **10**:153–172.

Hutchinson, W. G., S. M. Chilton, and J. Davis. 1995. Measuring non-use value of environmental goods using the contingent valuation method: problems of information and cognition and the application of cognitive questionnaire design methods. Journal of Agricultural Economics **46**:97–112.

Irwin, E. G. 2002. The effects of open space on residential property values. Land Economics **78**:465–480.

Isom, B. G. 1986. Historical review of Asiatic clam (Corbicula) invasion and biofouling of waters and industries in the Americas. American Malacological Bulletin, Special Edition **2**:1–5.

Jackson, J. B. C., K. E. Alexander, and E. Sala, editors. 2011. Shifting Baselines: The Past and the Future of Ocean Fisheries. Island Press, Washington, DC.

Jackson, L. E., J. C. Kurtz, and W. S. Fisher, editors. 2000. Evaluation Guidelines for Ecological Indicators. EPA/620/R-99/005. US Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

Jacobsen, J. B., J. H. Boiesen, B. J. Thorsen, and N. Strange. 2008. What's in a name? The use of quantitative measures versus "iconised" species when valuing biodiversity. Environmental and Resource Economics **39**:247–263.

Jakus, P. M., and W. D. Shaw. 2003. Perceived hazard and product choice: an application to recreational site choice. Journal of Risk and Uncertainty **26**:77–92.

Johnston, R. J. 2002. Combining economic and ecological indicators to prioritize salt marsh restoration actions. American Journal of Agricultural Economics **84**:1362–1370.

Johnston, R. J., E. Y. Besedin, E. Helm, and R. Stapler. 2013a. Overlaps and causality in aquatic ecosystem impacts of energy facilities: evaluating benefits of cooling water intake regulations under Section 316b of the Clean Water Act. Association of Environmental and Resource Economists Third Annual Summer Conference, Banff, Alberta, Canada.

Johnston, R. J., E. Y. Besedin, R. Iovanna, C. J. Miller, R. F. Wardwell, and M. H. Ranson. 2005a. Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: a meta-analysis. Canadian Journal of Agricultural Economics **53**:221–248.

Johnston, R. J., E. Y. Besedin, and R. F. Wardwell. 2003. Modeling relationships between use and nonuse values for surface water quality: a meta-analysis. Water Resources Research **39**:1363.

Johnston, R. J., and J. M. Duke. 2009. Willingness to pay for land preservation across states and jurisdictional scale: implications for benefit transfer. Land Economics **85**:217–237.

Johnston, R. J., J. J. Opaluch, M. J. Mazzotta, and G. Magnusson. 2005b. Who are resource nonusers and what can they tell us about nonuse values? decomposing user and nonuser willingness to pay for coastal wetland restoration. Water Resources Research **41**(7): W07017.

Johnston, R. J., and M. Ramachandran. 2014. Modeling spatial patchiness and hot spots in stated preference willingness to pay. Environmental and Resource Economics **59**:363–387.

Johnston, R. J., M. H. Ranson, E. Y. Besedin, and E. C. Helm. 2006. What determines willingness to pay per fish? A meta-analysis of recreational fishing values. Marine Resource Economics **21**:1–32.

Johnston, R. J., and R. S. Rosenberger. 2010. Methods, trends and controversies in contemporary benefit transfer. Journal of Economic Surveys **24**:479–510.

Johnston, R. J., and M. Russell. 2011. An operational structure for clarity in ecosystem service values. Ecological Economics **70**:2243–2249.

Johnston, R. J., E. T. Schultz, K. Segerson, E. Y. Besedin, and M. Ramachandran. 2012. Enhancing the content validity of stated preference valuation: the structure and function of ecological indicators. Land Economics **88**:102–120.

Johnston, R. J., E. T. Schultz, K. Segerson, E. Y. Besedin, and M. Ramachandran. 2013b. Stated preferences for intermediate versus final ecosystem services: disentangling willingness to pay for omitted outcomes. Agricultural and Resource Economics Review **42**:98–118.

Johnston, R. J., E. T. Schultz, K. Segerson, E. Y. Besedin, and M. Ramachandran. 2014. Guidelines for stated preference commodities: implications of biophysical production and utility theory. Fifth World Congress of Environmental and Resource Economists, Istanbul, Turkey, June 28–July 2.

Johnston, R. J., K. Segerson, E. T. Schultz, E. Y. Besedin, and M. Ramachandran. 2011. Indices of biotic integrity in stated preference valuation of aquatic ecosystem services. Ecological Economics **70**:1946–1956.

Johnston, R. J., S. K. Swallow, and D. M. Bauer. 2002. Spatial factors and stated preference values for public goods: considerations for rural land use. Land Economics **78**:481–500.

Johnston, R. J., T. F. Weaver, L. A. Smith, and S. K. Swallow. 1995. Contingent valuation focus groups: insights from ethnographic interview techniques. Agricultural and Resource Economics Review **24**:56–69.

Jordan, T., D. Weller, and D. Correll. 2003. Sources of nutrient inputs to the Patuxent River estuary. Estuaries **26**:226–243.

Jorgensen, S. L., S. B. Olsen, J. Ladenburg, L. Martinsen, S. R. Svenningsen, and B. Hasler. 2013. Spatially induced disparities in users' and non-users' WTP for water quality improvements: testing the effect of multiple substitutes and distance decay. Ecological Economics **92**:58–66.

Kaplowitz, M. D., F. Lupi, and J. P. Hoehn. 2004. Multiple methods for developing and evaluating a stated—choice questionnaire to value wetlands. Pages 503–524 *in* S. Presser, J. M. Rothget, M. P. Coputer, J. T. Lesser, E. Martin, J. Martin, and E. Singer, editors. Methods for Testing and Evaluating Survey Questionnaires. John Wiley and Sons, New York, NY.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries **6**:21–27.

Karr, J. R. 1991. Biological integrity: a long-neglected aspect of water resource management. Ecological Applications **1**:66–84.

Kenny, J. F., N. L. Barber, S. S. Hutson, K. S. Linsey, J. K. Lovelace, and M. A. Maupin. 2009. Estimated use of water in the United States in 2005. US Geological Survey, Reston, VA.

Kim, S.-I., and T. C. Haab. 2009. Temporal insensitivity of willingness to pay and implied discount rates. Resource and Energy Economics **31**:89–102.

Kling, C. L., D. J. Phaneuf, and J. Zhao. 2012. From Exxon to BP: has some number become better than no number? Journal of Economic Perspectives **26**:3–26.

Kontogianni, A., G. W. Luck, and M. Skourtos. 2010. Valuing ecosystem services on the basis of service-providing units: a potential approach to address the "endpoint problem" and improve stated preference methods. Ecological Economics **69**:1479–1487.

Kovacs, K. F., and D. M. Larson. 2008. Identifying individual discount rates and valuing public open space with stated-preference models. Land Economics **84**:209–224.

Krippendorff, K. 2013. Content Analysis: An Introduction to Its Methodology. 3rd ed. Sage Publications, Thousand Oaks, CA.

Ladle, R.J., and L. Gillson. 2009. The (im)balance of nature: a public perception time-lag? Public Understanding of Science **18**:229–242.

Landers, D. H., and A. M. Nahlik. 2013. A National Ecosystem Service Classification System for Final Ecosystem Goods and Services (NESCS). US Environmental Protection Agency, Office of Research and Development, Office of Research and Development, Western Ecology Division, Corvallis, OR.

Liebe, U., J. Meyerhoff, and V. Hartje. 2012. Test-retest reliability of choice experiments in environmental valuation. Environmental and Resource Economics **53**:389–407.

Lew, D. K., and K. Wallmo. 2011. External tests of scope and embedding in stated preference choice experiments: an application to endangered species valuation. Environmental and Resource Economics **48**:1–23.

List, J. A., and C. A. Gallet. 2001. What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics **20**:241–254.

Loomis, J. B. (2000). Vertically summing public good demand curves: an empirical comparison of economic versus political jurisdictions. Land Economics, 312-321.

Loomis, J. B. 2011. Incorporating distributional issues into benefit cost analysis: why, how, and two empirical examples using non-market valuation. Journal of Benefit-Cost Analysis **2**(01):1–24.

Loomis, J., P. Kent, L. Strange, K. Fausch, and A. Covich. 2000. Measuring the total economic value of restoring ecosystem services in an impaired river basin: results from a contingent valuation survey. Ecological Economics **33**:103–117.

Loomis, J. B., and R. S. Rosenberger. 2006. Reducing barriers in future benefit transfers: needed improvements in primary study design and reporting. Ecological Economics **60**:343–350.

Louviere, J. J., D. A. Hensher, and J. D. Swait. 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press, Cambridge, UK.

Luisetti, T., I. J. Bateman, and R. K. Turner. 2011. Testing the fundamental assumption of choice experiments: are values absolute or relative? Land Economics **87**:284–296.

Macmillan, D. C., and R. C. Ferrier. 1994. A bioeconomic model for estimating the benefits of acid rain abatement to salmon fishing: a case study in south west Scotland. Journal of Environmental Planning and Management **37**:131–144.

Mahan, B. L., S. Polasky, and R. M. Adams. 2000. Valuing urban wetlands: a property price approach. Land Economics **76**:100–113.

Martin-Ortega, J., and J. Berbel. 2010. Using multi-criteria analysis to explore non-market monetary values of water quality changes in the context of the Water Framework Directive. Science of the Total Environment **408**:3990–3997.

Massey, D. M., and G. Parsons. 2002. A random utility model of beach recreation. Pages 237-263 *in* N. Hanley, D. Shaw, and R. Wright, editors. The New Economics of Outdoor Recreation. Elgar, Cheltenham, UK.

McClelland, G. H., and W. D. Schulze. 1992. Methods for Measuring Non-use Values: A Contingent Valuation Study of Groundwater Cleanup. US Environmental Protrection Agency, Office of Policy, Planning and Evaluation.

McClelland, N. I. 1974. Water Quality Index Application in the Kansas River Basin. US Environmental Protection Agency–Region VII, Kansas City, MO.

McConnell, K. E., and W.-C. Tseng. 1999. Some preliminary evidence on sampling of alternatives with the random parameters logit. Marine Resource Economics **14**:317–332.

Meyer, A. 2013. Intertemporal valuation of river restoration. Environmental and Resource Economics **54**:41–61.

Miles, M. B., M. A. Huberman, and J. Saldaña. (2014). Qualitative Data Analysis: A Methods Sourcebook. 3rd ed. Sage Publications, Thousand Oaks, CA.

Milon, J. W., and D. Scrogin. 2006. Latent preferences and valuation of wetland ecosystem restoration. Ecological Economics **56**:162–175.

Mitchell, R. C., and R. T. Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the Future, Washington, DC.

Moeltner, K., and R. S. Rosenberger. 2008. Predicting resource policy outcomes via meta-regression: data space, model space, and the quest for "optimal scope." B E Journal of Economic Analysis and Policy **8**:31–31.

Morgan, D. L., and R. A. Krueger. 1998. The Focus Group Kit. 6 vols. Sage

Publications, Thousand Oaks, CA.

Morrison, M., J. Bennett, R. Blamey, and J. Louviere. 2002. Choice modeling and tests of benefit transfer. American Journal of Agricultural Economics **84**:161–170.

Mueller, J. M., and J. B. Loomis. 2008. Spatial dependence in hedonic property models: do different corrections for spatial dependence result in economically significant differences in estimated implicit prices? Journal of Agricultural and Resource Economics **33**:212–231.

Mueller-Dombois, D., and H. Ellenberg. 2002. Aims and Methods of Vegetation Ecology. Blackburn Press, Caldwell, NJ.

Murray, C., B. Sohngen, and L. Pendleton. 2001. Valuing water quality advisories and beach amenities in the Great Lakes. Water Resources Research **37**:2583–2590.

Nakano, D., and D. L. Strayer. 2014. Biofouling animals in fresh water: biology, impacts, and ecosystem engineering. Frontiers in Ecology and the Environment **12**:167–175.

Netusil, N., M. Kincaid, and H. Chang. 2014. Valuing water quality in urban watersheds: a comparative analysis of Johnson Creek Oregon and Burnt Bridge Creek, Washington. Water Resources Research **50**:4254–4268.

Niemeijer, D., and R. S. de Groot. 2008. A conceptual framework for selecting environmental indicator sets. Ecological Indicators **8**:14–25.

Niemi, G. J., and M. E. McDonald. 2004. Application of ecological indicators. Annual Review of Ecology, Evolution, and Systematics **35**:89–111.

Oliveira, J., M. T. Ferreira, P. Morgado, R. M. Hughes, M. T. Teixeira, A. Teixeira, R. M. Cortes, and J. H. Bochechas. 2009. A preliminary fishery quality index for Portuguese streams. North American Journal of Fisheries Management **29**:1466–1478.

Olsen, A. R., and D. V. Peck. 2008. Survey design and extent estimates for the Wadeable Streams Assessment. Journal of the North American Benthological Society **27**:822–836.

Opaluch, J. J., S. K. Swallow, T. Weaver, C. W. Wessells, and D. Wichelns. 1993. Evaluating impacts from noxious facilities: including public preferences in current siting mechanisms. Journal of Environmental Economics and Management **24**:41–59.

Parsons, G. R., D. M. Massey, and T. Tomasi. 1999. Familiar and favorite sites in a random utility model of beach recreation. Marine Resource Economics **14**:299–315.

Paterson, R. W., and K. J. Boyle. 2002. Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. Land Economics **78**:417–425.

Patton, M.Q. 2002. Qualitative Research and Evaluation Methods. 3rd ed. Sage Publications, Thousand Oaks, CA.

Peck, D. V., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, T. K. Magee, and M. Cappaert. 2006. Environmental Monitoring and Assessment Program - Surface Waters: Western Pilot Study: Field Operations Manual for Wadeable Streams. US Environmental Protection Agency, Washington, D.C. EPA/620/R-06/003.

Poor, P. J., K. J. Boyle, L. O. Taylor, and R. Bouchard. 2001. Objective versus subjective measures of water clarity in hedonic property value models. Land Economics **77**:482–493.

Powe, N. A. 2007. Redesigning Environmental Valuation: Mixing Methods within Stated Preference Techniques. Edward Elgar, Cheltenham, UK.

Reynolds and Rochon. 1991. Means-end based advertising research: Copy testing is not strategy assessment. Journal of Business Research, 22(2), 131-142.

Ringold, P. L., J. W. Boyd, D. H. Landers, and M. A. Weber. 2009. Report from the Workshop on Indicators of Final Ecosystem Services for Streams. US Environmental Protection Agency, Corvallis, OR.

Ringold, P. L., J. Boyd, D. Landers, and M. Weber. 2013. What data should we collect? A framework for identifying indicators of ecosystem contributions to human well-being. Frontiers in Ecology and the Environment **11**:98–105.

Ringold, P. L., J. W. Boyd, A. Nahlik, and D. Bernard. 2011. Report from the Workshop on Indicators of Rinal Ecosystem Services for Wetlands and Estuaries. US Environmental Protection Agency, Corvallis, OR.

Ringold, P. L., J. W. Boyd, M. Weber, and A. T. Herlihy. Forthcoming. Gaps in the capacity to describe stream ecosystem features that matter to people.

Ringold, P. L., M. Weber, and J. Boyd. 2012a. The Gap between What We Measure and What We Want to Know about Ecosystem Services in Streams and Their Linkages to Human Values. Ecological Society of America 2012 Annual Meeting, Portland, OR.

Ringold, P. L., M. Weber, J. Boyd, D. Landers, and A. Nahlik. 2012b. The Gap between What We Measure and What We Want to Know about Ecosystem Services in Streams and their Linkages to Human Values. ACES and Ecosystem Markets 2012, Fort Lauderdale, FL.

Rosenberger, R. S., and R. J. Johnston. 2009. Selection effects in meta-analysis and benefit transfer: avoiding unintended consequences. Land Economics **85**:410–428.

Rosenberger, R. S., and T. D. Stanley. 2006. Measurement, generalization, and publication: sources of error in benefit transfers and their management. Ecological Economics **60**:372–378.

Rubin, H. J., and I. S. Rubin. 2005. Qualitative Interviewing. 2nd ed. Sage Publications, Thousand Oaks, CA.

Rungie, C., R. Scarpa, and M. Thiene. 2014. The influence of individuals in forming collective household preferences for water quality. Journal of Environmental Economics and Management **68**:161–174.

Schaafsma, M., R. Brouwer, and J. Rose. 2012. Directional heterogeneity in WTP models for environmental valuation. Ecological Economics **79**:21–31.

Schiller, A., C. Hunsaker, M. A. Kane, A. K. Wolfe, V. H. Dale, G. W. Suter II, C. S. Russell, G. Pion, M. H. Jensen, and V. C. Konar. 2001. Communicating ecological indicators to decision makers and the public. Ecology and Society **5**:19.

Schkade, D. A., and J. W. Payne. 1994. How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation. Journal of Environmental Economics and Management **26**:88–109.

Schultz, E. T., R. J. Johnston, K. Segerson, and E. Y. Besedin. 2012. Integrating ecology and economics for restoration: using ecological indicators in valuation of ecosystem services. Restoration Ecology **20**:304–310.

Shannon, M. C., and C. M. Grieve. 1998. Tolerance of vegetable crops to salinity. Scientia Horticulturae **78**:5–38.

Shields, C. A., L. E. Band, N. Law, P. M. Groffman, S. S. Kaushal, K. Savvas, G. T. Fisher, and K. T. Belt. 2008. Streamflow distribution of non–point source nitrogen export from urban-rural catchments in the Chesapeake Bay watershed. Water Resources Research **44**:W09416.

Simpson, R. D. 1998. Economic analysis and ecosystems: some concepts and issues. Ecological Applications **8**:342–349.

Smith, V. K., C. Poulos, and H. Kim. 2002. Treating open space as an urban amenity. Resource and Energy Economics **24**:107–129.

Stephenson, W. 1993. Introduction to Q-methodology. Operant Subjectivity **17**(1): 1–13.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications **16**:1267–1276.

Stoddard, J. L., D. V. Peck, A. R. Olsen, D. P. Larsen, J. Van Sickle, C. P. Hawkins, R. M. Hughes, T. R. Whittier, G. Lomnicky, A. T. Herlihy, P. R. Kaufmann, S. A. Peterson, P. L. Ringold, S. G. Paulsen, and R. Blair. 2005. Environmental Monitoring and Assessment Program (EMAP): Western Streams and Rivers Statistical Summary. US Environmental Protection Agency, Office of Research and Development, Washington, DC.

Suter, G. W. II. 1993. A critique of ecosystem health concepts and indexes. Environmental Toxicology and Chemistry **12**:1533–1539.

Suter, G. W. II. 2001. Applicability of indicator monitoring to ecological risk assessment. Ecological Indicators **1**:101–112.

Swallow, S. K. 1996. Economic issues in ecosystem management: an introduction and overview. Agricultural and Resource Economics Review **25**:83–100.

Turner, R. K., S. Georgiou, and B. Fisher. 2008. Valuing Ecosystem Services: The Case of Multi-functional Wetlands. Earthscan, London, UK.

United Nations Environment Programme (UNEP). 2005. Concepts of Ecosystem Value and Valuation Approaches. Ecosystems and Human Well-being: A Framework for Assessment. UNEP, Nairobi, Kenya.

US Environmental Protection Agency. 2002. Estimation of National Economic Benefits Using the National Water Pollution Control Assessment Model to Evaluate Regulatory Options for Concentrated Animal Feeding Operations. EPA-821-R-03-009. Washington, DC.

US Environmental Protection Agency. 2008. EPA's 2008 Report on the Environment. Washington, DC.

US Environmental Protection Agency Office of Water. 2010. National Rivers and Streams Assessment.

US EPA Science Advisory Board. 2009. Valuing the Protection of Ecological Systems and Services: A Report of the EPA Science Advisory Board. EPA-SAB-09-012. US Environmental Protection Agency, Washington, DC.

Van Houtven, G., C. Mansfield, D. J. Phaneuf, R. von Haefen, B. Milstead, M. A. Kenney, and K. H. Reckhow. 2014. Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality. Ecological Economics **99**:40–52.

Van Houtven, G., J. Powers, and S. K. Pattanayak. 2007. Valuing water quality improvements in the United States using meta-analysis: is the glass half-full or half-empty for national policy analysis? Resource and Energy Economics **29**:206–228.

Vatn, A., and D.W. Bromley. 1994. Choices without prices without apologies. Journal of Environmental Economics and Management **26**(2):129–148.

Vaughan, W. J. 1986. The RFF water quality ladder. Appendix B in R. C. Mitchell and R. T. Carson, editors. The Use of Contingent Valuation Data for Benefit/Cost Analysis in Water Pollution Control: Final Report to the Environmental Protection Agency. Resources for the Future, Washington, DC.

Viscusi, W., J. Huber, and J. Bell. 2008. The economic value of water quality. Environmental and Resource Economics **41**:169–187.

Wallace, K. J. 2007. Classification of ecosystem services: problems and solutions. Biological Conservation **139**:235–246.

Weber, M., S. Caplan, and P.L. Ringold. 2014. Content Analysis to

Document Publicly Valued Ecosystem Services of Rivers and Streams [conference poster]. December. A Community on Ecosystem Services, Washington, DC. Available for download at http://www.conference.ifas.ufl.edu/aces/posters.html.

Weber, M., and P. L. Ringold. 2012. Ecosystems and People: Qualitative Insights. Environment, Health, and Society Bulletin 3(Spring): 2–8. http://www.epa.gov/ncer/ehs/bulletin/ehsspring2012/ehs_bulletin_spring2012.pdf.

Weber, M., and P. L. Ringold. 2014. Public Input on Stream Monitoring from Residents of the Willamette Valley, Oregon [conference poster]. December. A Community on Ecosystem Services, Washington, DC. Available for download at http://www.conference.ifas.ufl.edu/aces/posters.html.

Weber, M. A., and P. L. Ringold. 2015a. Priority river metrics for residents of an urbanized arid watershed. Landscape and Urban Planning **133**:37–52.

Weber, M. A., and P. L. Ringold. 2015b. Using Public Input to Define Ecosystem Services: Challenges and Insights. In preparation.

Weber, M.A. and S. Stewart. 2008. Public Valuation of River Restoration Strategies for the Middle Rio Grande. Restoration Ecology 17(6): 762-771.

Zandersen, M., M. Termansen, and F. S. Jensen. 2007. Testing benefits transfer of forest recreation values over a 20-year time horizon. Land Economics **83**:412–440.

Zhao, M., R. Johnston, and E. Schultz. 2013. What to value and how? Ecological indicator choices in stated preference valuation. Environmental and Resource Economics **56**:3–25.