Welfare, Wealth and Sustainability

Elena G. Irwin,[1] Sathya Gopalakrishnan,[2] and Alan Randall[3]

Forthcoming *Annual Review of Resource Economics*

This version: January 17, 2016

[1]Corresponding author: Department of Agricultural, Environmental and Development Economics, Ohio State University, 2120 Fyffe Road, Columbus, OH, USA 43210, irwin.78@osu.edu, 614-292-6449 (ph), 614-292-0078 (fx)

[2] Department of Agricultural, Environmental and Development Economics, Ohio State University, gopalakrishnan.27@osu.edu

[3] Department of Agricultural, Environmental and Development Economics, Ohio State University, randall.3@osu.edu

Table of Contents

Abstract

Growing concerns over climate change and the potential for large damages due to non-linear processes underscore the need for meaningful sustainability assessment of an economy. Economists have developed rigorous approaches to conceptualizing sustainability based on the paradigm of weak sustainability, which assumes infinite substitution between manufactured and natural capital stocks. In contrast, strong sustainability emphasizes physical limits to this substitution and the importance of maintaining the resilience of normally functioning biophysical processes. Recent progress in resource and environmental economics has demonstrated the feasibility of incorporating strong sustainability features, including tipping points, uncertainties and resilience, into welfare theoretic models to assess efficiency and optimal policies. Given that weak sustainability and intertemporal efficiency share the same concept of welfare, we ask: to what extent can these approaches be applied to evaluate sustainability? We highlight recent work on assessing sustainability in imperfect economies and dynamic models of intertemporal welfare that embed strong sustainability features.

Keywords: non-convexities, uncertainty, resilience, coupled human-natural systems, benefit cost analysis

Introduction

Welfare economics is fundamentally concerned with the implications of resource allocation for societal well-being. While the theory is sufficiently broad to consider both efficiency and equity concerns, applied welfare economics relies on the assumption of Benthamite welfare to justify benefit cost analysis (BCA), which aggregates benefits and costs without regard to distribution. An alternative framing of BCA appeals to contractarian reasoning: BCA identifies potential Pareto improvements (PPIs), i.e. changes that pass a hypothetical (Kaldor-Hicks) compensation test in which winners could compensate the losers. Given a well-developed and defensible set of methods for evaluating outcomes on the grounds of resource efficiency, economists have been slower to develop equivalent approaches for assessing distributional questions, most notably questions of inter-generational equity and the sustainability of economic systems. Because there is no assurance that an efficient economy is sustainable, a distinct welfare-theoretic approach to defining and operationalizing sustainability is necessary. An economic definition of sustainability emerged in the 1970s, spurred by growing public concerns over resource extraction and limits to growth: an economy is said to be weakly sustainable if welfare is non-declining over time. This approach is justified by an ethical view that emphasizes intergenerational equity—taking a Rawlsian approach, for example, intertemporal social welfare is maximized by maximizing the welfare of the least well-off generation (Solow 1974). Although motivated by different priorities, weak sustainability (WS) and intertemporal efficiency share the same basic concept of welfare. Economists have been able to readily adapt dynamic models of optimal resource consumption to evaluate the conditions under which the economy is weakly sustainable (e.g., Dasgupta and Heal 1979, Solow 1974, Hartwick 1977).

Since the 1970s, sustainability concerns, spurred by increasing awareness about climate change and its impacts, have shifted to questions about the Earth's capacity to absorb human impacts. New worries about tipping points, uncertainty over future states of the world and the risks of catastrophic outcomes have emerged, all of which stem from a fundamental belief that the Earth's capacity to absorb impacts is limited and that critical biophysical processes and ecosystem services must be preserved. Such assertions challenge the WS paradigm that presumes different forms of capital to be infinitely substitutable for each other—most notably, between natural and produced capital in production, and among goods and services in consumption. The implication is that, although relative prices may rise and in some cases precipitously, depletion of natural capital stocks does not undermine sustainability so long as sufficient investments in other types of capital are made. This worldview of generous substitutability is consistent with the modern experience of technical progress and increasing welfare with expanding substitution possibilities in both production and consumption. It is less consistent with a world in which tipping points and nonlinearities imply limits to the amount of environmental pollution or ecological damages that can be sustained without large declines in social welfare.

The notion of strong sustainability (SS) is motivated by these fundamental concerns, emphasizing limits to natural resources, limits to substitutability among natural and manufactured capital stocks, and limits to the resilience of normally functioning biophysical or ecosystem processes. SS suggests a moral imperative to restrain the consumption of critical natural capital (Ekins 2014), preserve unique and treasured environmental entities (Bishop 1978, Randall 2014), and pay greater attention to maintaining the resilience of ecosystems (Arrow et al. 1995). A recent example is the notion of planetary boundaries, defined as scientifically based levels of human perturbation of the Earth system beyond which the functioning of basic biophysical processes—e.g., climate change, biosphere integrity, land-system change, freshwater use, ocean acidification—may be substantially altered (Rockström 2009; Steffens et al. 2015). Different zones of human impacts have been identified for each of these biophysical processes, ranging from low-risk (a safe operating zone) to high-risk (beyond the so-called zone of uncertainty). The implication is that transgressing one or more of these boundaries creates substantial risk of destabilizing the current state of the Earth system, due to non-linear changes or tipping points, and that these changes may generate enormous societal impacts. However, the approach stops short of assessing trade-offs and instead advocates a safety-first SS approach that would restrict human impacts to a low-risk zone of influence.

Recent progress in resource and environmental economics has demonstrated that it is possible to incorporate key SS features, including tipping points and deep uncertainties, into welfare theoretic models and to apply BCA to assess implications for resource efficiency. Less attention has been given to explicitly assessing sustainability, however. This raises key questions: Do the familiar tools of intertemporal welfare economics, and more specifically, the methods of benefit cost analysis (BCA) applied in a dynamic setting, provide a suitable approach to evaluating WS? Can intertemporal models of resource consumption that account for ecological non-convexities, uncertainties, or other SS concerns be used to assess WS? How similar are the policy prescriptions of a WS approach that incorporate SS features versus an explicit SS approach that is based on identifying specific non-fungible critical resources and physical limits?

Here we consider these questions in the context of recent progress in the development of coupled human-natural systems models that incorporate one or more key SS features, but that have a welfare theoretic foundation and thus are capable of assessing WS. To set the stage, we begin with a summary of the earlier work in resource economics on sustainability,[1] including a discussion of efficiency versus sustainability and the conditions under which intertemporal utility maximization equates to WS and the practical evaluation of this criterion. A discussion of more recent work on imperfect economies follows, which illustrates that, although efficiency and sustainability are distinct concepts, the same wealth-based criterion for welfare analysis can be used in assessing both. We then take up recent developments in coupled human-natural systems models that have incorporated one or more elements of SS into utilitarian models, focusing particularly on uncertainty, tipping points, catastrophes, and non-convexities and examples from the nascent work in economics on resilience. The paper concludes with reflections on the foundational role of economics in assessing sustainability and some of the key challenges going forward.

Weak sustainability and welfare analysis

Assessing sustainability in perfectly competitive economies

In assessing the sustainability[2] of an economy, the role of the natural resource stock has led to different approaches: maintaining a steady state natural capital stock as a constraint in a dynamic welfare maximization framework, and non-declining welfare paths as an objective given an economy dependent on renewable and nonrenewable resource stocks (Heal 2000). The latter clearly defines the WS approach to sustainability and thus we focus our attention there. In 1974, Solow asked whether a society that uses exhaustible natural resources could nevertheless

---

[1] We only summarize the highlights of this work here. For a comprehensive review, see Pezzey and Toman (2002).

[2] Following Dasgupta (2001) and others, we use the terms "sustainability" and "sustainable development" interchangeably.

maintain human welfare indefinitely. Aggregate output or welfare at time t, Y_t, is determined as follows:

$$Y_t = e^{(a-d)t} f(N_t, K_t, L_t), \tag{1}$$

where N is the stock of natural resources, K is manufactured capital, L is labor, a is the rate of technological progress, and d is the population growth rate. Y, N, K and L are all indexes, such that particulars do not matter so long as the aggregates are maintained at levels that maintain welfare even as the composition of consumption changes over time. If (1) is taken to apply globally, people are assumed to be mobile, geographically, and occupationally; and the focus is on global rather than local natural resource and capital stocks. If N is exhaustible and $f(.)$ is Cobb-Douglas, the key result is that human welfare can be maintained for a very long time, so long as accumulation of K compensates fully for depletion of N, and a is as great as d (here we follow the common convention to assume that the $a \geq d$ condition will hold, and continue the discussion in per capita terms). This optimistic result depends crucially on generous substitutability, not just within the Y, N, K and L indexes but also between N and K, where the elasticity of factor substitution is constant and unitary. In effect, Solow granted special status to natural resources by assuming essentiality and exhaustibility, but immediately revokes it by assuming unlimited substitution of K for N. In the end, natural resources are nothing special in this model.

Hartwick (1977) showed that consumption is sustainable in a fixed technology economy with an essential exhaustible resource if net saving is everywhere zero (which requires that capital accumulation compensates exactly for resource depletion), the elasticity of substitution between resources and capital is unitary, and the elasticity of output with respect to capital is at least as great as the corresponding elasticity for the resource. The Hartwick rule would assure non-negative net saving in an exhaustible-resource-dependent economy by requiring that the scarcity rents from natural resource depletion be re-invested in reproducible capital. Substitutability between *N* and *K* is crucial for Hartwick sustainability: if the elasticity of substitution between produced capital and natural resources is less than one, then non-negative net saving cannot be maintained indefinitely (Dasgupta and Heal 1979, and Hamilton 1995) without an eventual decline in production and consumption.

This early work established the core proposition of WS and the central relationship between welfare and saving: if saving and investment, including investment in human capital, are insufficient in each period to compensate for resource depletion and environmental degradation, welfare eventually must decline. Thus WS is not about how much we consume, but about whether we save and invest enough to compensate for our consumption. The central role of saving and investment in WS suggests a strong link between WS and wealth.

A seminal paper by Weitzman (1976) provides another cornerstone, demonstrating that given a first best setting and with additional assumptions, e.g., constant technology and linear utility, a comprehensive measure of net national product (NNP) expressed in utility terms is proportionate to the maximized present discounted value of intertemporal social welfare. The key insight is that future welfare prospects, which are inherently unobservable, are indicated by the current value of changes in all capital assets in the economy, broadly defined to include natural, human, and social forms of capital in addition to manufactured capital. This theoretical equivalence has spawned a vast literature, both theoretical and applied, seeking to extend and apply the result to a test of weak sustainability. Asheim and Weitzman (2001) show that changes in wealth, measured as the growth in real net national product (where prices are deflated by a

Divisia index of consumption prices) indicates the change in welfare in the economy. The implication is that negative levels of net saving reduce utility in future periods. Pezzey and Toman (2002, pp. 184 – 185) and Pezzey (2004) use the model developed by Asheim and Weitzman to show that genuine saving (GS, an ideal account of net savings in a resource-using economy) provides a one-sided sustainability test in the Hartwick tradition – with negative GS, there must be negative welfare growth at any instant.[3,4] However, the result by Asheim and Weitzman (2001) relies on the specialized properties of the Divisia price index. As both Aronsson et al. (2004) and Dasgupta (2009) show, a static equivalent to future welfare does exist under more general conditions, but it is equal to the sum of NNP and the consumer surplus associated with the current level of consumption. This eliminates the strict proportionality between NNP and future welfare and, from a practical viewpoint, makes it much more difficult to implement a sustainability test using this framework, given the obvious challenges of estimating the consumer surpluses associated with all consumption goods.

Weak sustainability, wealth, and efficiency

The link between WS and wealth, central as it is, introduces a nagging issue: WS sustains welfare for each generation whereas wealth is the present-value of welfare summed over the generations. Many economists argue that the apparatus of applied welfare economics – the PPI

---

[3] Pezzey and Burke (2014, Appendix A) provide a clear statement of exactly what is assumed in order to show that GS provides a one-sided test of WS.

[4] Pezzey and Toman (2002) show that the Hartwick rule cannot offer an exact policy prescription for sustainability in the real world, because observed prices are not generated by an underlying sustainability objective function (to put it another way, sustainability prices can be observed only once sustainability has been achieved).

criterion and BCA – applied in an intertemporal setting is consistent with WS in most respects (e.g. Pezzey and Toman, 2002; Dietz and Neumayer, 2007; Pearce et al., 2006). Dissecting this argument, we endorse the view that WS and the PPI criterion share the same concept of contemporaneous welfare, but may diverge in a multigenerational context—the defining aspect of WS—in that WS sustains welfare for each generation while the PPI-BCA approach maximizes the sum of welfare over time, i.e. maximizes wealth. This introduces two potential divergences from WS. First, the PPI criterion would countenance intergenerational fluctuations in welfare so long as the multigenerational sum of welfare was maximized. This concern that could be addressed readily, at least in principle, via intergenerational redistribution, as suggested by Bishop (1993) and Stavins et al. (2003).

Second, the PPI criterion counts only the present value of future benefits and costs. Solow (1974) was quite clear that welfare in WS is not discounted – in other words, that WS implies an ethical commitment to intergenerational equality, which clearly distinguishes WS from intertemporal optimization. With discounting as commonly practiced, the present value of welfare over time is likely to be maximized by choosing a declining consumption path over time, a property that has troubled some economists for more than fifty years (Koopmans 1960). We argue that the two approaches can be reconciled readily, at least in principle, with help from Ramsey's (1928) equation $\rho_t = \delta + \eta \cdot g_t$, where ρ is the discount rate, δ is the rate of time preference, g is the rate of growth in welfare, η is marginal utility of consumption, and t is time. Assuming η is approximately equal to 1, $\rho_t = \delta + g_t$, then the discount rate has two components: the rate of time preference and the growth rate of welfare. In a series of papers, Asheim and various colleagues (e.g. Asheim *et al* 2012, Dietz & Asheim 2012, and Zuber & Asheim 2012) have argued that intergenerational equality requires discounting future growth in welfare – for example, discounting at the rate *g* in the case where welfare is expected to grow smoothly at *g* – but not for time preference. Imposing positive time preference in the process of discounting would be contrary to the intergenerational equity ethic inherent in WS and invoked by Solow.

Economists tend to assume that discount rates with neutral time preference would still be positive, reflecting productivity and scarcity of capital, just as interest rates would be positive in a weakly sustainable economy. But that would hold only if we expect the future to be better off than the present, an assumption that some have called into question based on concerns about the prospects of tipping points and catastrophes. An inescapable circularity arises when considering the right discount rate in this setting: we can only act rationally toward the future if we know how much better or worse off than us future generations will be (Asheim *et al* 2012, Dietz & Asheim 2012, and Zuber & Asheim 2012).

It is readily apparent that any economy, no matter how (in)efficient, can sustain some welfare level greater than zero, including very low levels that clearly would be undesirable. It can be argued that a sustainability goal is incomplete without specifying a minimum acceptable level of welfare. Bishop (1993) and later Stavins et al. (2003) address this by suggesting the following modification: dynamic efficiency is a necessary, but insufficient condition for weak sustainability. There are many candidate policies or actions to change the intertemporal allocation of resources that would generate positive net benefits, as measured in present value terms. Not all would lead to non-declining welfare over time and, if the discount rate includes positive time preference, welfare will eventually decline. Stavins et al. appeal to the idea of potential compensation, arguing that a dynamically efficient economy has the potential to be made sustainable via appropriate intergenerational transfers.

While this approach is appealing to economists who think in efficiency terms, the Achilles heel of potential compensation—that hypothetical compensation is a kind of hypothesis, not a kind of compensation—applies with even more force to potential intergenerational compensation. A reasonable fear is that redefining WS to elevate dynamic efficiency and downplay intergenerational equity risks kicking the sustainability can further down a potentially very long intergenerational road. Efficiency is not the same as sustainability. Both are good, but only one of them, efficiency, has the potential to be advanced by laissez-faire markets. This suggests an important continuing role for WS policies.

Assessing sustainability in imperfect economies

The correspondence between current wealth and optimized intertemporal welfare relies on a first best world in which the economy evolves along an optimal consumption path.[5] As Dasgupta (2001, 2009), Arrow, Dasgupta, Maler (2003) and others (e.g., Aronsson et al. 2004) have argued, to assume that the economy is perfectly competitive and at its full optimum may be a useful starting point, but ultimately requires a heroic set of assumptions. This is particularly relevant for a theory that seeks to account for natural capital stocks, given the pervasiveness of missing markets for ecosystem services.

---

[5] In deriving his canonical result, Weitzman (1976) assumes that population and technology are constant and that markets are perfectly competitive. These assumptions ensure that the economy is at a full optimum and that the economic system is autonomous, i.e., that time itself does not have a direct influence on utility or profits. Therefore, it is possible to draw a correspondence between a static measure of the economy and the optimal discounted value of all future utilities. If these assumptions are relaxed, then the NNP welfare measures contain unobservable forward looking terms that cannot be measured with observable data.

Dasgupta (2001, 2009) and coauthors (Dasgupta and Mäler 2000, Arrow et al. 2003) relax the assumption of convex production, a necessary assumption for achieving an optimal allocation in a decentralized economy,[6] and develop a framework for assessing sustainable development in imperfect economies. The approach is similar to the wealth-based welfare theoretic approach of the earlier literature, but does not assume that the government seeks to optimize intertemporal social welfare. The analysis starts with a standard Ramsey-Koopmans form of intergenerational welfare, defined as $V(t) = \int_t^\infty \left[U(C(\tau))e^{-\delta(\tau-t)}\right]d\tau$, where $C(t)$ is the rate of aggregate consumption, U is per period social utility and δ is the social discount rate. Aggregate production is assumed to depend on multiple capital stocks, whose current levels determine the state of the economy in period t, $S(t) = \left(K(t), L(t), N(t)\right)$, where $K(t)$ is manufactured capital, $L(t)$ is human capital and $N(t)$ is natural capital. The productivity of $N(t)$ is assumed to be characterized by a minimum threshold, below which economic output would eventually decline to zero. This is a highly stylized representation, but nonetheless captures the strong sustainability principle of finite carrying capacity or planetary boundaries that, if transgressed, can lead to non-marginal declines in resources and services.

A non-optimizing "resource allocation mechanism" is assumed to be known and reflects institutional constraints that may arise from the structure of property rights, tax rates or allocation of public goods. This mechanism provides a many-to-one mapping of capital stocks into an economic program that determines the current and future allocation of these stocks and flows. The mapping is not assumed to lead to an optimal, or even an efficient, economy. Assuming a given resource allocation mechanism, the intertemporal welfare function can be

---

[6] This is established by the Second Theorem of Welfare Economics.

written as an explicit function of initial stocks in period t, $V(S(t))$,[7] the value function. Assume that shadow prices are defined as the partial derivative of the social welfare function with respect to each of the stocks.[8] Then differentiating V with respect to t provides an exact measure of a local change in intergenerational welfare over time as the change in comprehensive wealth or equivalently, comprehensive investment, $I(t)$:

$$\frac{dV(S(t))}{dt} = p(t)\frac{dK(t)}{dt} + q(t)\frac{dL(t)}{dt} + n(t)\frac{dN(t)}{dt} = I(t), \qquad (2)$$

where $p(t)$, $q(t)$, and $n(t)$ are the shadow or accounting prices for the respective capital stocks. From this, it follows that comprehensive wealth corresponds to a linear index of well-being, which equates to the aggregate value of all productive capital stocks in the economy evaluated at constant shadow prices,

$$W(t) = p(t)K(t) + q(t)L(t) + n(t)N(t). \qquad (3)$$

Evaluated at a point in time, Equation (2) provides a local test of weak sustainability: the economy is sustainable in time t if $I(t) \geq 0$. In order to assess whether the economy is weakly sustainable over a longer time period, it is necessary to also account for capital gains on the assets that have accrued over the time interval by deducting the aggregate value of these gains from the difference in comprehensive wealth between the two dates, $t = 0$ and $t = T$:

---

[7] Arrow et al. (2003) and Dasgupta (2009) also consider the case in which V depends directly on t and therefore is non-autonomous, which substantially complicates the welfare index.

[8] To address the problem of defining shadow price at the point of a discontinuity, given nonconvex production, Arrow et al. (2003) and Dasgupta (2009) assume that it can be defined within a sufficiently small neighborhood of this point, e.g., assuming V is continuous on either side of the discontinuity.

$$V(T) - V(0) = W(T) - W(0) - G(0, T), \tag{4}$$

where G is the aggregate value of capital gains between the two periods.

For a small change in capital assets, Dasgupta (2001, 2009) and Arrow et al. (2003) demonstrate the correspondence between the change in comprehensive wealth that results and changes in the present discounted sum of utilities over time arising from the corresponding future changes in consumption:

$$V\big(S(t) + \Delta S(t)\big) - V\big(S(t)\big) = \int_t^\infty U'(C(\tau))\Delta C(\tau)e^{-\delta(t-\tau)}\, d\tau, \tag{5}$$

where $\Delta S(t)$ is a small change in capital assets. This suggests a criterion for intertemporal BCA. For example, suppose a project uses labor, manufactured capital and resources as inputs and produces manufactured capital output, which may be consumed today or in a future period. The project results in a change in capital assets over the time span of the project, but generates changes in consumption over all future periods. The equivalence of changes in wealth over a given time span to the discounted sum of utilities from all future changes in consumption due to the project implies that comprehensive investment is an appropriate welfare criterion for project evaluation. However, this correspondence holds only for sufficiently small changes in the economy, in which it is reasonable to assume constant shadow prices and no change in consumer surpluses.

Wealth accounting and shadow prices

In principle, an ideal system of national accounts would tell society whether wealth is non-declining over time and thus whether it is weakly sustainable. Pearce and Atkinson (1993) were among the first to draw a practical linkage between changes in current wealth and the implications for changes in future welfare and a test of weak sustainability. Building on the early resource economics models of an optimizing economy and conditions for WS laid out by the Hartwick rule, they implement a WS test by estimating an adjusted savings rate of countries that accounted for resource depletion and environmental pollution. This measure, now called Genuine Saving (GS), can be defined as the sum of the changes in stocks at in time t of each of the various kinds of natural and produced capital, each weighted by its virtual price:

$$GS_t = \sum p_{it} \Delta K_{it} \tag{6}$$

where $i = (1, \ldots, n)$ is an exhaustive list of the various forms of capital with virtual prices p_i. K_i includes gross national saving, net investment in human capital, depreciation, depletion of minerals and energy, net depletion of forests, net depletion of water resources in terms of quantity and quality), depletion of biodiversity, net pollution damage (including damage from greenhouse gases), and net degradation of soil. The World Bank developed this approach further (e.g., Hamilton and Clements 1999) and has developed the ANS (adjusted net savings) accounts that measure GS for more than 200 countries, although it fair to say that investment in human capital is proxied poorly by education expenditures and non-marketed components seem to be undervalued. Similarly, their Comprehensive Wealth accounts are more credible for tangible capital than for human, social, and institutional capital, which are represented as the otherwise unexplained residual in the estimated relationship between wealth and output, and usually amount to more than half the total wealth of rich countries.

Arrow et al. (2012) use the per capital growth rate of comprehensive wealth, adjusted for total factor productivity growth, to implement a WS test. Although the approach derives from a different theoretical foundation, namely that of imperfect economies, it is similar in spirit to the World Bank's Comprehensive Wealth accounts and expands on the set of assets that are valued.

Some key differences emerge in the estimation of the capital assets and their values, however. For example, Arrow et al. calculate the value of health capital using estimates of the value of a statistical life, which generates estimates that are much larger than the estimated values of any of the other assets (produced, natural, and human) and much larger than the "intangible residual" wealth estimated by the World Bank that includes health capital among other forms of unobserved capital.

A primary empirical challenge is the calculation of shadow prices, or accounting prices as Arrow et al. (2012) and others refer to them. The welfare basis for the comprehensive wealth-based approach rests clearly on the assumption that these prices reflect the full discounted social value of an incremental change in a given capital asset. Thus, a forecast of how the economy is likely to evolve over time is needed, which requires a dynamic model of the economy, natural system and the potential linkages between the human and natural systems. In addition, because the economy is not assumed to be on an optimal path, the marginal rates of substitution and transformation are not equal, implying that obtaining a welfare-based measure of wealth requires demand-side measurement of willingness-to-pay measures. While market prices may reasonably reflect the social marginal value of some capital stocks, this will not be the case for many others and in particular, many types of natural capital stocks subject to externalities or in some cases imperfect competition. Arrow et al. (2012) incorporate capital depreciation, e.g., from carbon emissions, and non-market forest benefits into their estimation of wealth changes, but otherwise fall short of this ideal. With few exceptions, their approach to natural capital valuation calculates shadow price as the market price net extraction costs.

Fenichel and Abbott (2014) provide an important methodological contribution to natural capital valuation that could bridge a critical gap between the theory of welfare evaluation in imperfect economies and empirical implementation. Following Dasgupta (2009), the current value Hamiltonian for a non-optimized, autonomous system is shown to be the current return on the present value of all net benefits over time, i.e., $\delta V(t)$ where δ is the discount rate and $V(t) = \int_t^\infty \left[W(s(\tau))e^{-\delta(\tau-t)}\right]d\tau$, W is an index of net social benefits at time t and $s(t)$ is the natural capital stock. This fundamental result is used to derive an expression for a natural capital asset price, $\partial V/ds$, equal to the marginal ecosystem service flows, W_s, adjusted by the scarcity effects (capital gains or losses) resulting from an additional unit of natural capital, $\dot{p}$, divided by the discount rate adjusted for the effect of an additional unit of natural capital on the overall growth rate of the capital stock, $\dot{s}$:

$$\frac{\partial V}{\partial s} = p = \frac{W_s(s, x(s)) + \dot{p}}{\delta - \dot{s}} \tag{8}$$

where $\dot{s} = G_s(s) - f_s(s, x(s))$, $G(s)$ is the ecological production function, and $f(s, x(s))$ is the damage function associated with the human impacts of consuming $s(t)$. Consumption is determined by a non-optimal feedback rule $x(s)$—or what Dasgupta and others call the economic program—that reflects human activities and decisions that directly or indirectly impact $s(t)$, e.g., management policy that directly determines resource consumption, as well as institutional constraints, social norms or other types of non-optimizing behaviors that may indirectly influence $s(t)$. The marginal expressions G_s and f_s are respectively the marginal productivity of natural capital in the absence of humans, and the marginal damages from human impacts given an increase in the natural capital stock. The “effective” discount rate (i.e., denominator), and therefore price, increases with the rate of damages from human impacts and decreases with the ecological productivity of the natural capital asset.

Fenichel and Abbott (2014) apply this approach to the Gulf of Mexico reef fish complex, using harvesting and price data, previous studies (Zhang and Smith 2011) to parameterize the fish stock growth function and fishing effort response equation, and focus on the market value of the fishery. They use numerical methods to solve for the value of $\dot{p}$, conditional on $s(t)$, and demonstrate the advantages over simpler capital valuation methods that do not fully account for the price effects of ecological and human conditions that determine stock levels and changes over time. In focusing only on market net benefits, their estimated price falls short of what is ideally needed for wealth accounting—that is, the full marginal social value of the fish stock. Doing so is possible using the same approach, but requires the additional steps of estimating the full range of producer surpluses, e.g., including that of the fish processors, and the full consumer surpluses generated by fish stocks, including both the market benefits from final goods consumption and non-market values of ecosystem services that generate net benefits beyond those accruing to the fishery.

Addressing strong sustainability concerns

That the well-developed apparatus of applied welfare analysis in combination with capital-theoretic models of dynamic resources can be usefully applied to evaluate WS is encouraging. What remains to be considered is whether and how the concerns of strong sustainability—most notably, uncertainty, non-convexities and resilience—can be incorporated and the welfare implications of these SS features.

The mainstreaming of strong sustainability: catastrophes, fat tails, and ambiguity

The sustainability question was motivated for economists by Malthusian worries: business as usual generation after generation may eventually exhaust the carrying capacity of the planet. Developments in WS were mostly congenial to mainstream economists, quibbles with discounting notwithstanding. Nevertheless, a distinct minority of economists conceded a role for SS in maintaining a safe minimum standard of conservation as a prudential precaution and to preserve unique and valued environmental assets (Bishop 1978, Farmer and Randall 1998).

More recently, climate science has raised the possibility of catastrophic global warming, which would seem to call for a different kind of sustainability discussion. WS and BCA are designed to help resolve resource allocation problems in an ordinary world, and they can seem out of their depth when the prospect of global catastrophe is taken seriously. A vigorous discussion among economists has arisen in the last fifteen years. Chichilnisky (2000) shows that the expected utility approach is insensitive to unlikely but potentially catastrophic events, and proposes a ranking criterion for potential projects or policies that gives weight to expected utility considerations and to the desire to avoid unlikely but catastrophic outcomes. Barrieu and Sinclair-Desgagné (2006) offer a rather elaborate weighting formulation in an attempt to deal simultaneously with two key precautionary principle issues: scientific uncertainty and risk-benefit trade-offs. The scales are tilted toward harm-avoidance in two ways: more harmful scenarios are over-weighted and the benefits-harm trade-off is tilted in favor of harm-avoidance. Gollier *et al.* (2000) and Gollier and Treich (2003) have used the concept of scientific uncertainty, a real options approach to managing the wait for better scientific information, and a Bayesian approach to learning from new information to identify conditions under which precautionary measures are efficient. Gollier-Treich precaution is of a circumstantial kind that can be revised in response to emerging information.

Among economists and the informed public, discussion of response to catastrophic prospects was intensified by the Stern-Nordhaus debates on climate change in the mid-2000s

(Stern 2007, Nordhaus 2007). Climate change presents an overwhelming challenge to business-as-usual resource allocation tools: a complex narrative that poses the highly uncertain prospect of a meta-catastrophe (a catastrophe that triggers a suite of catastrophes) decades and centuries into the future, driven by greenhouse gas (GHG) emissions that persist in the atmosphere, and preventable only at considerable cost to current and near-future generations. Standard BCA thinking seems unable to endorse climate mitigation action strong enough to meet the broadly-agreed target of no more than 1.5 to 2 degrees Celsius increase in global temperature. Some economists are inclined to say “end of story,” while others have wondered in various constructive ways whether the problem might be attributed more to inadequacies in BCA thinking than to a genuine lack of urgency in climate change mitigation. For example, Fisher and Le (2014) identify a consensus among integrated assessment modelers that potential impacts of climate change are not large enough to warrant aggressive mitigation efforts in the near term, and argue that this view is misleading for three reasons: projections end typically at the year 2100, but the greatest impacts are expected later than that; the models focus on most likely harm, paying little attention to worst-cases; and the discount rates used are inappropriate for serious threats to long-term global welfare.

In this section, we address a clear and urgent question: how do we make investment decisions to reduce the possible, if unlikely, prospect of catastrophe, and does the BCA framework comprehend the question well enough to provide a reasonable answer?

Initial discussion focused on the discount rate, which reflects two distinct concerns: the rate of time preference and the productivity of capital. As the debate about time preference seemed stalemated, Weitzman (2009) offered a game-changing argument: observing that an array of predictions from reputable climate models (Meinshausen et al. 2009) suggested a nontrivial possibility (around 6 to 10 percent) of 6 degrees Celsius warming, he suggested that we should be more attentive to the potentially catastrophic tails of the distribution. He noted that fat-tailed distributions were consistent with complex-systems properties (Roe and Baker 2007). Various objections can be entertained—statistical reasoning is inapplicable because the data are not observations but predictions, and they are not independent because different climate models share much of the same structure and data (Millner and Dietz 2013); and we assume too much in taking these probabilities seriously given the level of uncertainty that surrounds them (Gilboa *et al* 2009). But nonetheless, the essential point holds: concentration on the most likely outcome is misplaced when catastrophe is more than trivially probable. Weitzman offers a "dismal theorem" to the effect that WTP to avert catastrophe may approach infinity.

Weitzman's argument moved the focus toward the way we think about risk, generally understood as chance of harm. The standard analysis of risk treats chance as arising because outcomes are generated by known stochastic processes, and frequently invokes the analogy of well-specified games of chance. But there are at least two other kinds of chance (Randall 2011): chance arising from our lack of understanding of the process that generates outcomes—an insight that leads, for example, to consideration of uncertainty, ambiguity, and gross ignorance, and to Bayesian learning formulations; and chance that arises because the system that generates outcomes is itself changing, as might occur in complex adaptive systems.

The term "ambiguity" arises in a variety of guises. Millner, Dietz, and Heal (2013) and Dietz and Gollier (2015) follow Klibanoff et al (2009), who define ambiguity as a lack of confidence in "soft" evidence with respect to probability of harm, leading the decision maker to apply more than ordinary caution.[9] They obtain a general comparative static result that, given explicit restrictions on the relevant functions, optimal abatement increases with ambiguity aversion. Then they introduce scientific ambiguity into the well-known DICE model of the climate-economy system and conclude that, under plausible assumptions, ambiguity increases WTP to avert catastrophic outcomes to a much greater extent than does ordinary risk.

Groping to better understand the idea of "unknown unknowns", which goes well beyond ambiguity as formulated by Klibanoff *et al* (2009) and Izhakian (2015), Henri and Henri (2002) analyze the implications of gross ignorance, and Schipper (2014) suggests that the emerging literature on unawareness eventually may provide a more satisfactory analysis of unknown unknowns.

Traeger (2014), addressing both ambiguity and discounting, finds that when uncertainty about the climate system is considered, the interaction and correlation between economic growth and project payoffs becomes a major ingredient for evaluating climate change and pricing carbon. A relatively small inter-temporally correlated risk suffices to cut the discount rate back to pure time preference, eliminating the growth effect in discounting. Given our argument above that very-long-run time preference should approach zero, a zero allowance for growth would suggest a zero discount rate.

Showing renewed interest in an older question, Martin and Pindyck (2015) ask how the benefits and costs of catastrophe avoidance might be affected by the concurrent threat of several

---

[9] Other definitions of ambiguity are narrower, e.g. Izhakian's (2015), which defines uncertainty as having two components, risk and ambiguity; and broader to include ambiguity in specifying magnitudes and even kinds of outcomes, as well as their likelihood.

independent kinds of catastrophes. Not surprisingly, they conclude that even if all the catastrophe-prevention projects pass a BCA test one by one, it is generally not justified to implement them all – the package may not pass the BCA test and, even if it does, some of the components will fail a BCA test given that the others already are part of the package – a special case of the more general Hoehn and Randall (1989) theorems concerning a multi-project agenda. They offer an *ad hoc* method of setting priorities in an “all projects pass the BC test one by one, but not as package” context.

It should be noted that, while climate change is one of the potential catastrophes considered by Martin and Pindyck, it is really a different kind of threat. Whatever climate change occurs will be with us a very long time. In contrast, many of the competing catastrophes—floods, nuclear terrorism, bioterrorism, maybe even super-viruses—are episodic events and their incidence and magnitude may be increased by excessive warming. In that reading of the reality, a climate catastrophe is a meta-catastrophe, and the urgency of preventing it may be diminished less by the presence of additional catastrophic threats.

Nonconvex production

Uncertainty over the future path of resource consumption stems in large part from our scientific lack of understanding of the complex dynamics that coupled human natural systems sometimes exhibit and the difficulty in incorporating these features into our models. A fundamental result in welfare economics – that equilibrium market prices can result in efficient allocation of resources – rests on the assumption that all possible transformation possibilities between goods and services form a convex set. The determination of optimal natural capital investment and consumption paths to accomplish sustainable welfare goals is further complicated by potential non-convexities in human and natural systems (Dasgupta and Mäler

2003). Non-convexity leads to the possibility of multiple stable states and the long run steady state towards which the system evolves often depends on the interactions between human decisions and the natural (or ecological) system.

Resource economists have made significant advances in the optimal management of natural resources with non-convex ecological production functions represented in the state transition function (Starrett 1972, Tahvonen and Salo 1996, Fenichel et al. 2015). For example, in the problem of examining the optimal path of nutrient runoff into a shallow lake that provides local services (the "shallow lake" problem), the state transition dynamics representing the accumulation of phosphorous in the lake is given by a single state transition function (Brock and Starrett 2003; Mäler et al. 2003). The state dynamics in this problem are generally represented by: $\dot{x}(t) = a(t) - bx(t) + \frac{x^2(t)}{x^2(t)+1}$, where the rate of accumulation of phosphorous stock, $x(t)$, in the lake depends on additional loading, $a(t)$, but also displays hysteresis. In these systems the stock transition dynamics depend not only on the current loading but also on the total accumulated stock at time t. $bx(t)$ represents the natural rate of absorption of phosphorous. Non-convexity in the state dynamics can lead to multiple equilibriums. In some cases the transition to a eutrophic state is irreversible (Fig 1) and avoiding eutrophic state in the future requires reductions in loading in earlier time periods. If the lake reaches a eutrophic stable state, even a complete reduction of loading to zero does not enable the system to recover. The evolution of the system into an oligotrophic or eutrophic equilibrium depends on both the initial level of phosphorous stock in the lake and the rate of loadings over time.

Another common approach to describe non-convexities in bioeconomic models is to introduce depensation in the dynamics of a renewable natural capital stock, where the per capita growth rate declines as a stock gets smaller. When there exists a critical level of the stock below which the growth rate becomes negative, this is called critical depensation. Such representation of the system with a single state variable that reflects nonlinearities and complex ecological interactions provide insight for policy makers (Barbier et al. 2008). However, this approach presents greater challenge when there are multiple capital stocks with nonlinear interactions that affect both current and future stream of aggregate welfare flow. Non-convex production possibilities have also been examined in the context of economic growth and development. Models that examine nonlinear interactions between nutrition and human productivity have been important in explaining the prevalence of poverty traps in developing economies (Dasgupta and Ray 1987; Dasgupta 1997).

Systems with multiple stable states that display hysteresis or path dependence, where the state transition function depends not only on the current state but also on the history of previous states, may be easily depicted in one dimension but are more challenging to conceptualize in higher dimensional systems. The challenge for policy design to manage non-convex systems is that the natural capital system can evolve to undesirable stable states that limit the possibility of reaching a desirable state in the future. In these complex systems with multiple steady states, transition from one steady state can often represent a regime shift to an alternative state that has different characteristics and responsiveness to perturbations. This suggests the need for forward-looking adjustments in the short run to avoid undesirable outcomes. If production and consumption decisions are made without adjusting for nonlinear dynamics and the shadow value of natural capital stocks, then welfare maximizing allocations can push society in a direction that is qualitatively worse off, beyond inefficiency in consumption and allocation. The possibility of unanticipated regime shifts could suggest an optimal consumption path that results in lower aggregate social welfare when we consider qualitative changes in the stock of ecological or natural capital. The presence of ecological non-convexities then implies that investments or disinvestments in natural capital may be irreversible (Dixit and Pindyck 1994). These irreversibilities indicate possible tipping points that have significant effects on society's wealth and sustainability.

Ecological thresholds and tipping points are important in understanding complex systems that are jointly determined by biophysical dynamics and economic decisions. In systems with multiple stable states, the cumulative effect of economic decisions (natural capital extraction, or pollution for example) can push the system away from a stable attractor, resulting in a dramatic shift in the possible set of future actions. The moment marking this transition to instability is often termed a 'tipping point' (Poston and Stewart 1977). A change in system stability could be marked by a subtle and smooth transition to a new attractor. However, the loss of system stability can also remain hidden for some time until observable change is triggered by a sudden perturbation to the system, such as a major natural hazard or a policy change.

Coupled models of human and natural systems are essential to better understand nonlinear feedbacks that could lead to emergent patterns that are qualitatively different from those derived from examining the biophysical or the economic dynamics in isolation (Murray et al. 2011). Tipping points and regime shifts that mark a change in the ecological system dynamics have been examined using bioeconomic models that set up the post tipping point outcome as a new bioeconomic problem and simultaneously solving the infinite horizon problem (Homans and Horie 2011; Horan et al 2011). Advances in optimal control models that incorporate multiple thresholds that could result in a catastrophic loss of natural capital stock, have derived conditions under which optimal regulation paths (of CO_2) can drive the probability of crossing one or more thresholds to zero (Naevdal 2003, 2006; Naevdal and Oppenheimer 2007) and reduce a complex

N-dimensional risk structure into a simpler problem. With the possibility of endogenous regime shifts that could potentially result in the collapse of ecological stocks, optimal policy requires investments to maintain larger natural capital stocks (Polasky et al. 2011).

Recent advances in integrated assessment models are beginning to examine tipping points and allow for endogenous interactions between abrupt irreversible climate shifts, social welfare, and optimal policy responses. When policy makers utilize new information to learn and update their beliefs about the possibility of reaching a tipping point, numerical models show that different types of climate shifts that can occur in the post-tipping regime can result in optimal policy paths are qualitatively different before the tipping point is reached (Lemoine and Traeger 2014). The introduction of stochastic risks of approaching a tipping point following a climate shock in an integrated assessment model shows that optimal climate policy paths are considerably different from the outcomes in a deterministic model. These integrated climate-economy models further shows that when the probability of reaching a tipping point is endogenously determined by the policy path chosen in the pre-tipping period, optimal policy adjustments involve more stringent regulation (a carbon tax in most cases) that can reduce the probability (Lemoine and Traeger 2014) and can reduce the impact or potential damages of a climate shift. (Jensen and Traeger 2014; Cai et al. 2015; Lontzek et al. 2012). For example, Cai et al. (2015) show that a 5 percent annual probability of a 5 percent loss in welfare, occurring with at 4°C increase of the global surface temperature, would increase the optimal carbon tax more than threefold. The possibility of crossing critical thresholds that can shift the future dynamics of natural capital stocks and the stream of welfare flow essentially increase the shadow value of natural capital.

In essence, the presence of ecological non-convexities and possible regime shifts, driven by climate forcing, indicate that the optimal path of consumption and investment in natural capital derived in a dynamic welfare maximizing framework could mimic a SS rule, applied at the threshold. Furthermore, in an imperfect economy with non-convex natural capital stocks that display hysteresis, a resource allocation rule that does not require an inter-temporal welfare-maximizing criterion could lead to a desirable long run outcome (Arrow et al. 2003).

Resilience as a capital asset

The potentially large costs that can come from a regime shift suggest that the system's ability to absorb shocks and thereby avoid a tipping point may have value. Resilience can be defined as the capacity of a system to remain within a given regime, i.e., the domain of attraction of a given stable attractor, in systems where multiple equilibria are possible (Holling 1973, 1996). The greater the resiliency of a system, the larger the shock it can absorb without undergoing a regime shift. Mäler (2008) argues that resilience can be conceived of as a stock that protects the desired state of the system by reducing the risk of undesirable changes that would result from a regime shift. Like any capital stock, as the resilience of the system to an undesirable shift grows scarcer, it becomes more valuable. However, like many ecosystem services, resilience is a public good—it is valuable to each user of an ecosystem, but no individual has the right incentive to optimally invest in it.

Accounting for resilience requires a dynamic model of a capital resource stock whose contribution to the resilience of the system to a specific threshold can be quantified. For example, Chen et al. (2009, 2012) develop a regional migration model with endogenous lake amenities. A phase plot diagram is used to quantify resilience, which is a function of the stocks of human population and lake pollution, as the distance of any given state of the system to the boundary that separates the domains of attraction of two stable attractors and illustrate the trade-offs between resilience and efficiency.

Mäler (2008), Mäler and Li (2010) develop a capital theoretic approach to valuing resilience. They define the shadow value of resilience, the presented discounted marginal benefits of an additional unit of resilience, as the difference in the social welfare associated with the good versus bad state, weighted by the change in probability of the bad state occurring, summed and discounted over time. Walker et al. (2010) apply this method to valuing the resilience of the Goulburn-Broken Catchment (GBC) in South East Australia, one of the country's most important agricultural regions, to a salinity threshold. They use historical data on salinization and soil fertility to identify the threshold and measure resilience of the system to a high saline regime by the distance from the water table to the threshold. They estimate the probability that the system will shift from the non-saline to the saline regime as a function of this distance measure and use market prices to translate this probability into value terms.

Efforts to model resilience are in their infancy and face many challenges. It is tempting nonetheless to speculate on how measures of resilience could be useful. The value of resilience could be incorporated into wealth accounting and used to assess trade-offs among resilience and other capital assets, thus offering a means of determining the socially desirable levels of natural (and other) capital stocks that provide resilience. Valuation of resilience could also be used to assess the net benefits of policies that can increase incentives towards resilience enhancing activities, particularly those that vary dynamically within systems, and the social costs of a loss in resilience. Presumably the incorporation of resilience into BCA and WS analyses would increase investments in capital stocks that contribute to resilience and in so doing, mimic a SS rule that would impose direct physical limits to maintain critical natural capital stocks.

Conclusions

The impact of human actions on natural resources and the environment and our ability to sustain welfare given the Earth's physical limits has concerned classical economists since the 18th century. Early thinkers like Malthus and Jevons raised concerns about the sustainability of welfare paths in the wake of industrial revolution (Barnett and Morse 1963; Jevons 1865). The development of neo classical welfare theory shifted focus to optimal allocation decisions on the basis of estimating changes in consumer and producer surpluses given a price or quantity change, while advances in resource economics focused on the central idea of nature as capital. The link between welfare-theoretic surplus measures and the time evolution of capital stocks that are the inputs into production and consumption have only recently gained attention from resource and environmental economists. However, the integration of these two distinct approaches is essential both for modeling coupled human-natural systems and for valuing natural capital stocks (Fenichel and Abbott 2014).

As we have emphasized in this paper, the framework used to define sustainability depends critically on assumptions about substitution in production and consumption across different forms of capital, and in particular between natural and manufactured capital. The distinction and potential inconsistency between WS and SS arguments stem largely from fundamental differences in the manner in which economists and natural scientists view this substitution. Economists following welfare theoretic BCA are concerned with substitution at the margin whereas ecologists are typically concerned with the limits to substitutability at critical thresholds (Fenichel and Zhao 2014). The incorporation of SS concerns into dynamic models of intertemporal welfare maximization begins to bridge this gap**.** Applying a sustainability objective that does not impose optimality conditions could further bridge the WS and SS gap by permitting greater flexibility in modeling resource allocation decisions and integrating complex interactions among capital stocks.

Another disciplinary difference is that economists rely heavily on the assumption of convexity to derive optimality conditions whereas natural scientists focus on nonlinear interactions in complex biophysical systems. To understand the peculiarities of natural capital, namely non-convexities in how it is produced, and the potential nonlinear feedbacks that can arise from coupled human-natural systems, it is important to increase the economic and ecological realism used to represent the dependence of intertemporal welfare on dynamic capital stocks. This requires collaborative efforts to advance empirical analyses that measure the shadow value of natural capital stocks and mathematical techniques to model these complex systems.

Wealth-based measures of welfare provide the means of incorporating more realistic dynamics into intertemporal welfare models, but rely critically on shadow prices to translate changes in capital assets into changes in welfare. Calculating shadow prices requires a forecast of the future evolution of capital stocks, which assumes knowledge of the salient human behaviors, environmental processes, and their interactions and feedbacks over time and space. Thus, in attempting to evaluate sustainability, we come to the same conundrum that others have encountered: assessing whether or not we are sustainable in the future requires that we know the economic program and are able to forecast the future, something that is inherently unknowable, on this basis.

But what are the alternatives? We can ignore the future, pursue approaches that don’t rely on shadow prices, or develop value-based approaches that account for a long-run that reeks of unpredictability. The former brings us back to traditional welfare assessment, e.g., estimating WTP measures given observed or hypothetical changes ecosystem services without linking this to the underlying stocks or production. This limits welfare analysis to static BCA and thus misses the core sustainability question.

The second approach is essentially SS, which involves a commitment to sustain specific capital stocks at specified levels. But what particulars should be sustained, and why? If some natural capital stocks really are critical, utilitarians can support SS criteria with respect to those forms of capital based on the welfare implications of not maintaining these critical levels. But SS criteria might be adopted also by people who attribute intrinsic value to particular natural entities, and those with a deep ecology perspective, to mention just two of many ethical stances that might come into play. Should all things natural enjoy SS protections because they are natural, or should SS protections hinge on moral claims specific to particular entities? In utilitarian formulations, the moral claims would be not just specific but contingent on preferences and circumstances that determine the relative trade-offs.

For these reasons, most economists interested in sustainability assessment are likely to side with the last approach: value-based approaches that rely on dynamic modeling of capital stocks that can assess trade-offs and changes in welfare over the long run, but that also can account for the deep uncertainties that arise—both from our limited knowledge of systems as well as their complex nature that causes them to adapt and change over time. Such an approach does not lack for challenges, but nonetheless we see promise and progress and have attempted to highlight these here.

Acknowledgements

We thank Nicholas Irwin for his outstanding research assistance and gratefully acknowledge support from National Science Foundation Grants CNH 1114934, WSC 1058059, DEB 0410336.

Literature cited

Aronsson T, Karl-Gustaf L, Backlund K. 2004. *Welfare Measurement in Imperfect Markets: A Growth Theoretical Approach*. Cheltenham, UK: Edward Elgar Publishing

Arrow KJ, Bolin B, Costanza R, Dasgupta P, Folke C, Holling CS, Jansson BO, Levin S, Mäler KG, Perrings C, Pimentel D. 1995. Economic growth, carrying capacity, and the environment. *Science* 268:520-21

Arrow KJ, Dasgupta P, Mäler KG. 2003. Evaluating projects and assessing sustainable development in imperfect economies. *Environmental and Resource Economics* 26(4): 647-685

Arrow KJ, Dasgupta, P, Goulder, LH, Mumford KJ, and Oleson, K. 2012. Sustainability and the measurement of wealth. *Environment and development economics* 17(3):317-353

Asheim GB, Mitra, T, Tungodden, B. 2012. Justifying sustainability. *Economic Theory* 49(2):267-292

Asheim, GB, Weitzman, ML. 2001. Does NNP growth indicate welfare improvement? *Economics Letters* 73(2): 233-39

Barbier EB, Koch EW, Silliman BR, Hacker SD, Wolanski E, Primavera J, Granek EF, Polasky S, Aswani S, Cramer LA, Stoms D, Kennedy CJ, Bael D, Kappel CV, Perillo GME, Reed DJ. 2008. Coastal ecosystem–based management with nonlinear ecological functions and values. *Science* (319):321-323

Barnett HJ, Morse C. 1963. *Scarcity and growth: The economics of natural resource availability*. The Johns Hopkins Press, Baltimore

Barrieu P, Sinclair-Desgagné, B. 2006. On precautionary policies. *Management science, 52*(8): 1145-1154

Bishop RC. 1978. Endangered species and uncertainty: the economics of a safe minimum standard. *American journal of agricultural economics*, *60*(1):10-18

Bishop RC. 1993. Economic efficiency, sustainability, and biodiversity. *Ambio* 22(3-4):69-73

Brock WA, Starrett D. 2003. Managing systems with non-convex positive feedback. *Environmental and Resource Economics* 26(4):575-602

Cai Y, Judd KL, Lontzek TS. 2016. DSICE: A dynamic stochastic integrated model of climate and economy. *RDCEP Working Paper* No. 12-02

Cai Y, Judd KL, LontzekS T, Naritae D. 2015. Environmental tipping points significantly affect the cost−benefit assessment of climate policies. *Proceedings of the National Academy of Science* 112(15):4606-4611

Chen Y, Irwin EG, Jayaprakash C. 2009. Dynamic modeling of environmental amenity-driven migration with ecological feedbacks. *Ecological Economics* 68(10):2498-2510

Chen Y, Jayaprakash C, Irwin EG. 2012. Threshold management in a coupled economic–ecological system. *Journal of Environmental Economics and Management* 64(3):442-455

Chichilnisky G. 2000. An axiomatic approach to choice under uncertainty with catastrophic risks. *Resource and Energy Economics*, *22*(3):.221-231

Dasgupta P. 1997. Nutritional status, the capacity for work, and poverty traps. *Journal of Econometrics* 77(1):5-37

Dasgupta P. 2001. Valuing objects and evaluating policies in imperfect economies. *Economic Journal* 111(471):1-29

Dasgupta P. 2009. The welfare economic theory of green national accounts. *Environmental and Resource Economics* 42(1):3-38

Dasgupta P, Heal G. 1979. *Economic Theory and Exhaustible Resources*. Cambridge, UK: Cambridge University Press

Dasgupta P, Mäler K.G. 2000. Net national product, wealth, and social well-being. *Environment and development economics* 5(1):69-93

Dasgupta P, Mäler K.G. 2003. The economics of non-convex ecosystems: Introduction. *Environmental and Resource Economics* 26(4):499-525

Dasgupta P, Ray D. 1987. Inequality as a determinant of malnutrition and unemployment: policy. *The Economic Journal* 97(385):177

Dietz S, Gollier C, Kessler L. 2015. The climate beta. *The Grantham Institute Working paper 190*

Dietz S, Asheim GB. 2012, Climate policy under sustainable discounted utilitarianism. *Journal of Environmental Economics and Management* 63(3):321-335

Dietz S, Neumayer E. 2007. Weak and strong sustainability in the SEEA: Concepts and measurement. *Ecological Economics* 60(4):617-636.

Dixit AK, Pindyck RS. 1994. *Investment under uncertainty*. Princeton, NJ: Princeton University Press

Ekins P. 2014. Strong sustainability and critical natural capital. In Atkinson G, Dietz S, Neumayer E, and Agarwala M (eds) *Handbook of Sustainable Development, 2nd Ed*. Cheltenham, UK: Edward Elgar 55-71

Farmer MC, Randall A. 1998. The rationality of a safe minimum standard. *Land Economics* 1: 287-302

Fenichel EP, Abbott JK. 2014. Natural capital: from metaphor to measurement. *Journal of the Association of Environmental and Resource Economists* 1(1):1-27

Fenichel E, Gopalakrishnan S, Bayasgalan, O. 2015. 7. Bioeconomics: nature as capital. *Handbook on the Economics of Natural Resources*:165

Fisher AC, Le PV. 2014. Climate policy: science, economics, and extremes. *Rev Environ Econ Policy* 8(2):307-327

Gilboa I, Postlewaite A, Schmeidler D. 2009. Is it always rational to satisfy savage's axioms? *Economics and Philosophy* 25:285-296

Gollier C, Jullien, B, Treich N. 2000. Scientific progress and irreversibility: an economic interpretation of the 'Precautionary Principle'.*Journal of Public Economics*, *75*(2): 229-253

Gollier C, Treich N. 2003. Decision-making under scientific uncertainty: the economics of the precautionary principle. *Journal of Risk and Uncertainty*, *27*(1), pp.77-103

Hamilton K. 1995. Sustainable development, the Hartwick rule and optimal growth. *Environment and Development* 5(4): 393-411

Hamilton K, Clemens M. 1999. Genuine savings rates in developing countries. *The World Bank Economic Review* 13(2):333-356

Hamilton K, Hartwick J.M. 2005. Investing exhaustible resource rents and the path of consumption. *Canadian Journal of Economics* 38(2):615–2.

Hamilton K, Ruta G. 2009. Wealth accounting, exhaustible resources and social welfare. *Environmental and Resource Economics* 42(1):53-64

Hartwick J. 1977. Intergenerational equity and the investing of rents from exhaustible resources. *American Economic Review* 67(5):972-974

Heal G. 2000. *Valuing the future: economic theory and sustainability*. Columbia University Press

Henry C, Henry, M. 2002. *Formalization and Applications of the Precautionary Principle.* Université Catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES) Discussion Paper 2002009.

Hoehn JP, Randall A. 1989. Too many proposals pass the benefit cost test. *American Economic Review* 79:544-551

Holling CS. 1973. Resilience and stability of ecological systems. *Annual review of ecology and systematics* 4(1):1-23

Holling CS. 1996. Surprise for science, resilience for ecosystems, and incentives for people. *Ecological Applications* 6(3):733-735

Homans F, Horie, T. 2011. Optimal detection strategies for an established invasive pest. *Ecological Economics* 70:1129-1138

Horan R, Fenichel E, Drury K, Lodge D. 2011. Managing ecological thresholds in coupled environmental–human systems. *Proceedings of the National Academy of Sciences* 108:7333-7338

Izhakian Y. 2015. *A theoretical foundation of ambiguity measurement* New York.

Jensen S, Traeger, C.P. 2014. Optimal climate change mitigation under long-term growth uncertainty: Stochastic integrated assessment and analytic findings. *European Economic Review* 69:104-125

Jevons WS. 1865. *The coal question: An inquiry concerning the progress of the nation, and the probable exhaustion of the coal-mines*. Macmillan

Klibanoff P, Marinacci M, Mukerji S. 2005. A smooth model of decision making under ambiguity. *Econometrica* 73(6):1849–1892

Koopmans TC. 1960. Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society* 1:287-309

Lemoine D, Traeger CP. 2014. Watch your step: Optimal policy in a tipping climate. *American Economic Journal: Economic Policy* 6(1):137-166

Lontzek TS, Cai Y, Judd KL, Lenton T.M. 2015. Stochastic integrated assessment of climate tipping points indicates the need for strict climate policy. *Nature Climate Change* 5(5):441-444

Mäler KG. 2008. Sustainable development and resilience in ecosystems. *Environmental and Resource Economics* 39(1):17-24

Mäler KG, Li C.Z. 2010. Measuring sustainability under regime shift uncertainty: a resilience pricing approach. *Environment and Development Economics* 15(6):707-719

Mäler KG, Xepapadeas A, De Zeeuw, A. 2003. The economics of shallow lakes. *Environmental and Resource Economics* 26(4):603-624

Margolis M, Naevdal E. 2008. Safe minimum standards in dynamic resource problems: conditions for living on the edge of risk. *Environmental and Resource Economics* 40(3):401-423

Martin IWR, Pindyck RS. 2015. Averting catastrophes: The strange economics of Scylla and Charybdis. *American Economic Review* 105(10):2947–2985

Meinshausen M, Meinshausen N, Hare W, Raper SCB, Frieler K, Knutti R, Frame DJ, Allen MR. 2009. Greenhouse-gas emission targets for limiting global warming to 2C. *Nature* 458(7242):1158.

Millner A, Dietz S, Heal G. 2013. Scientific ambiguity and climate policy. *Environmental and Resource Economics* 55(1):21–46

Murray A, Gopalakrishnan S, McNamara D, Smith M. 2013. Progress in coupling models of human and coastal landscape change. *Computers & Geosciences* 53:30-38

Naevdal E. 2003. Optimal regulation of natural resources in the presence of irreversible threshold effects. *Natural Resource Modeling* 16(3):305-333

Nævdal E. 2006. Dynamic optimization in the presence of threshold effects when the location of the threshold is uncertain–with an application to a possible disintegration of the Western Antarctic Ice Sheet. *Journal of Economic Dynamics and Control* 30(7):1131-1158

Nævdal E, Oppenheimer M. 2007. The economics of the thermohaline circulation—A problem with multiple thresholds of unknown locations. *Resource and energy economics* 29(4):262-283

Nordhaus W. 2007. A review of the Stern review on the economics of climate change. *Journal of Economic Literature* 45:686-702

Pearce D, Atkinson G. 1993. Capital theory and the measurement of sustainable development: an indicator of weak sustainability. *Ecological Economics* 8(2):103-108

Pearce D, Atkinson G, Mourato S. 2006. *Cost-benefit analysis and the environment: recent developments*. Paris: OECD Publishing

Pezzey JC. 2004. One-sided sustainability tests with amenities, and changes in technology, trade and population. *Journal of Environmental Economics and Management*, *48*(1):613-631

Pezzey JC, Burke P. 2014. Towards a more inclusive and precautionary indicator of global sustainability. *Ecological Economics* 106:141–154

Pezzey JC, Toman M. 2002. Progress and problems in the economics of sustainability. In Tietenberg T, and Folmer H. (eds.) *International Yearbook of Environmental and Resource Economics 2002/2003*. Cheltenham, UK, and Brookfield, VT: Edward Elgar 165-232

Polasky S, De Zeeuw A, Wagener F. 2011. Optimal management with potential regime shifts. *Journal of Environmental Economics and Management* 62(2):229-240

Poston T, Stewart I. 1977. *Catastrophe theory for scientists and engineers*. San Francisco: Dover Publications

Ramsey FP. 1928. A mathematical theory of saving. *The Economic Journal* 1:543-559

Randall A. 2014. Weak sustainability, conservation, and precaution. In Atkinson G, Dietz S, Neumayer E, and Agarwala M (eds) *Handbook of Sustainable Development*, 2nd Ed. Cheltenham, UK: Edward Elgar 160-172

Randall A. 2011. *Risk and Precaution*. Cambridge, UK and New York, NY: Cambridge University Press

Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, Lenton TM, Scheffer M, Folke C, Schellnhuber H, Nykvist B. 2009. A safe operating space for humanity. *Nature* 461(7263):472-475

Roe G, Baker M. 2007. Why is climate sensitivity so unpredictable? *Science* 318(5850):629–632

Schipper B. 2014. Unawareness—A gentle introduction to both the literature and the special issue. *Mathematical Social Sciences* 70:1-9

Solow R. 1974. Intergenerational equity and exhaustible resources. *Review of Economic Studies, Symposium on the Economics of Exhaustible Resources* 29-46

Starrett DA. 1972. Fundamental nonconvexities in the theory of externalities. *Journal of Economic Theory* 4(2):180-199

Stavins RN, Wagner AF, Wagner G. 2003. Interpreting sustainability in economic terms: dynamic efficiency plus intergenerational equity. *Economics Letters* 79(3):339-343

Steffen W, Richardson K, Rockström J, Cornell SE, Fetzer I, Bennett EM, Biggs R, Carpenter SR, de Vries W, de Wit CA, Folke C. 2015. Planetary boundaries: Guiding human development on a changing planet. *Science* 347(6223):1259855

Stern N. 2007. *The Economics of Climate Change*. New York, NY: Cambridge University Press

Tahvonen O, Salo S. 1996. Nonconvexities in optimal pollution accumulation. *Journal of Environmental Economics and Management* 31(2):160-177

Traeger CP. 2014. Why uncertainty matters - discounting under intertemporal risk aversion and ambiguity. *Economic Theory* 56(3):627-664

Walker B, Pearson L, Harris M, Maler KG, Li CZ, Biggs R, Baynes T. 2010. Incorporating resilience in the assessment of inclusive wealth: an example from South East Australia. *Environmental and Resource Economics* 45(2):183-202

Weitzman ML. 1976. On the welfare significance of national product in a dynamic economy. *The Quarterly Journal of Economics* 90(1):156-162

Weitzman ML. 2009. On modeling and interpreting the economics of catastrophic climate change. *Review of Economics and Statistics* 91(1): 1-19

Zhang J, Smith MD. 2011. Estimation of a generalized fishery model: A two-stage approach. *Review of Economics and Statistics*, *93*(2), pp.690-699

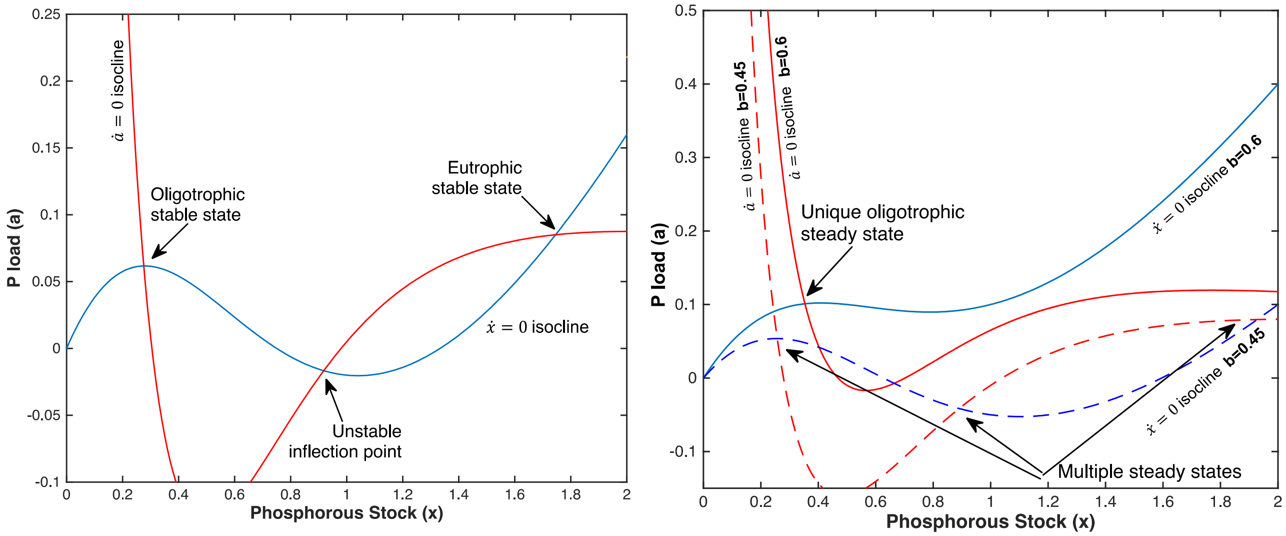
Zuber S, Asheim GB. 2012. Justifying social discounting: the rank-discounted utilitarian approach. *Journal of Economic Theory* 147(4):1572-1601

A note on population

(sidebar: please place in section “Weak sustainability and welfare analysis”)

Clearly, population matters to any serious discussion of sustainability across generations, yet it is often discussed only briefly and then reserved for future work due to the difficulties it introduces. Population is endogenous in the long run and not readily predictable. In appealing to Solow-type models it has become conventional to note that, so long as $a \geq d$ and the necessary substitutability conditions are satisfied, then welfare per capita can be sustained—and then proceed as though $a \geq d$ were assured. Some recent articles have addressed population more directly (e.g. Dietz and Asheim 2012), but these remain the exception. Furthermore, the endogeneity of population greatly complicates intergenerational ethics, too, a point emphasized by Rawls (1971) when explaining his hesitance in expanding his analysis to the intergenerational context. It is not at all clear how a smaller population sustaining a higher level of welfare is to be valued ethically relative to a larger population sustaining a somewhat lower level of welfare. In intergenerational ethics it has become common to frame each generation as a generational-person, a fiction that advances the discussion by avoiding the population question, much as the $a \geq d$ assumption does in economic formulations of WS.

Figure 1



(a) Phase plane representing phosphorous accumulation in a shallow lake. Natural rate of P absorption is low (b = 0.48). If the stock exceeds the unstable inflection point, the system will move into a eutrophic state even if load is reduced to zero.

(b) Change in climate forcing that reduces the absorptive capacity of the lake (changing b from 0.6 to 0.45) can move the system from a unique stable oligotrophic state to one with multiple stable states.

Figure 1: Phase plane representing phosphorous accumulation in a shallow lake.