Optimal climate policy including animal welfare, and methods for quantifying animal wellbeing in policy and decision analyses, and estimating optimal tradeoffs against human wellbeing[*]

Mark Budolfson (Rutgers) and Dean Spears (Texas)

Version 2.0

1. Introduction

Economists, philosophers, and other researchers have long recognized that the wellbeing of animals is a neglected but potentially important part of the correct evaluation of decision-making in both the public and private sphere (Fleurbaey et al. 2018, McShane 2018, Ng 2016, Sunstein climate, Cowen 2006, Jamieson 2008, Sunstein and Nussbaum 2005, Ng 1995, Singer 1975). This is because, like humans, animals also experience different levels of wellbeing depending on decisions made by others – for example, decisions about natural resource use, animal agriculture, etc. As a result, mainstream decision evaluation approaches that take maximizing wellbeing (perhaps subject to various constraints) to be the goal of decision analysis must include accounting for the wellbeing of animals in order to fully account for the consequences of decisions for wellbeing.

However, animal wellbeing has been ignored in most decision analysis models in the past, primarily for two reasons. First, there is often a bias against giving any weight to the wellbeing of animals in decision analysis. However, many researchers believe that this bias does not have any theoretical justification, and instead is an unjustifiable bias of the same sort found in colonialist decision-making of an earlier era that gave no weight to the wellbeing of people who lived as subjects of the colonial powers (or, similarly, decision-making in an earlier era that systematically ignored the wellbeing of women) (Sunstein and Nussbaum 2006, Singer 1975).

In what follows, we set aside this important issue of whether there is any justification for giving less weight to the wellbeing of animals. Instead, we focus on removing a second barrier to incorporating animal welfare that has vexed even those who fully endorse the idea that the wellbeing of animals should be given

---

[*] Thanks to 2016 Montreal Animals conference, 2016 Graz Climate Economics Conference, 2017 Texas Environmental Conference, 2017 Pacific APA ISEE, 2018 Stockholm Institute for Future Studies Conference on Population and Demography, 2018 Santa Fe Workshop on Economics and Philosophy.

weight in decision analysis. This is the problem of how to estimate animal wellbeing and make tradeoffs between human and animal wellbeing – in short, the problem of interspecies comparisons. This reason for ignoring animal wellbeing is a matter of an insufficiently developed theory and practice of animal wellbeing and its relation to human wellbeing. Our goal in this paper is develop pragmatic methods that may not resolve all theoretical puzzles, but would be sufficient in many contexts to remove this second barrier to including animal welfare in decision analyses.

2. Background: The anthropocentric bias in decision analysis models, and the need for sensitivity tests in Decision Analysis Models given normative uncertainty

Decision analysis models (DAMs) are used to evaluate alternative possible courses of action, and to select which action to take. These models have three defining elements:

1. *Impact assessment*: estimation or assumption of the likely consequences/outcomes of the alternative courses of action. (E.g. a model of the consequences for temperature, or for energy use, of one course of action vs. other possible courses of action.)

2. *Valuation* of these consequences/outcomes. (For example, assign dollar values, or utility to consequences in 1.)

3. *A decision objective*: A clear formula for choosing between courses of action that have differently valued consequences/outcomes. Sometimes (but not always) this includes a representation of uncertainty over the possible outcomes of each course of action. (For example, given the valuation in 2, the decision objective might be to choose the course of action that maximizes GDP, or maximizes utility, or (with uncertainty represented) maximizes expected utility.)

The most influential models that inform public policy are DAMs in this sense. DAMs are often (and increasingly) used in other contexts beyond public policy, such as by NGOs deciding what programs to fund, individuals deciding what charities to give to, and any other context where a decision must be made about a complex problem that can be modeled, and where valuation metrics can be designed to represent better and worse possible outcomes.

DAMs are especially pervasive in evaluations of environmental policy, especially when environmental outcomes are coupled with human systems, and especially when there are complex feedbacks between various elements of these systems. A leading example is in connection with climate change, where complex DAMs dominate the analyses. Climate DAMs that issue policy recommendations include a sophisticated impact assessment module, and a representation of the feedbacks between human activity and environmental outcomes. At the same time, much simpler DAMs are used for many policy decisions – environmental and otherwise, as any cost-benefit policy analysis has the form of a DAM, often with a simple decision theory such as 'prefer policy options with higher net benefits in dollar terms' – which, when taken to its logical conclusion, this implies maximizing GDP as the policy objective.

In this paper we focus on a method for improving valuation of animal wellbeing in DAMs. Many researchers from across disciplines, and many citizens, believe that an important problem with the standard valuation methodology in these models is that valuation is always *anthropocentric* valuation (Schmidtz and Shahar 2018, Sandler 2018, McShane 2018, Palmer et al. 2014, Sarkar 2012, Jamieson 2008, O'Neill et al. 2008, Ng 1995, Singer 1975). In other words, valuation is always in terms of the *value of outcomes to humans only*, and thus assigns no intrinsic value to the wellbeing of non-human animals. For example, in standard models, the deaths of millions of birds due to climate change has disvalue only insofar as the deaths of those birds have disvalue to humans. But many would object that this way of valuing animal lives is fundamentally incorrect because it ignores any intrinsic value of the birds' own utility irrespective of additional utility they may have to humans. (Similarly, leading DAMs assign no intrinsic value to the health of ecosystems, which is a separate criticism of DAMs that we set aside here (Chan et al. 2016, Eyal and Schlenker 2016, Dasgupta 2014, Alcamo et al. 2003, Dasgupta 2001, Costanza et al. 1997).)

So, according to critics of anthropocentric DAMs, what is needed is the addition of valuation of non-human animals. That is the project we take up in this paper. In what follows, we outline a theoretically-well-grounded and empirically-based method for adding intrinsic (non-anthropocentric) valuation of the wellbeing of animals to DAMs.

A complication is that there is widespread disagreement about how to assign value to non-human animals even among those who believe their wellbeing should be assigned intrinsic value. And many other researchers and citizens, led by many mainstream economists, would altogether reject the idea of assigning any non-anthropocentric value to animals. In light of this disagreement, one might wonder whether there is any 'objective' way to make progress in decision analyses along the lines that we pursue here.

Nonetheless, there is a straightforward and established method for representing alternative but contested normative frameworks within DAMs – namely, the method of adding a sensitivity test that investigates how the decisions that are recommend by the DAM would change if alternative values were assigned to key normative parameters instead, where this sensitivity test is calibrated to the range of values that are widely defended by citizens and scholars. A leading example of this method of adding a sensitivity test in the context of normative uncertainty and disagreement is provided by the 'discounting debate' in climate change policy, in which there is fundamental and irresolvable disagreement over the normative question of whether to discount the wellbeing of future persons simply because they exist in the future (Nordhaus 2007, Dasgupta 2007, Weitzman 2007, Stern 2006). In this similar context of normative uncertainty, it is now considered best practice to test the sensitivity of policy recommendations in climate DAMs to a range of different discount rates that represent the range of values for those parameters that are widely defended (USIIWGSCC, Nordhaus 2013). This method of adding a sensitivity test in the context of disagreement about normative parameters is also illustrated by other examples, including investigating the sensitivity of optimal policy to alternative assumptions about 'population ethics' (e.g. whether policy should attempt to maximize the average level of wellbeing in society, or the total sum of wellbeing in society) (Arrhenius forthcoming, Scovronick et al. 2017, Dasgupta 2001, Parfit 1984), and contexts in which there is disagreement about empirical parameters (Gillingham at al. 2018), including those that are normatively significant (Nordhaus 2007, Dennig et al. 2015).

Because this method of adding a sensitivity test is available, and because without it existing DAMs impose a single normative framework upon society that many would reasonably reject, it is arguably essential for such a sensitivity test or some other substitute to be added to these models insofar as they are used to make public policy and other decisions that have important effects on society (Scanlon 1998, Rawls 1972).

Thus, the methods proposed below will be used to generate a standard sensitivity test with different reasonable principled valuations that can be added to existing DAMs, which will range from the current status quo of 'no non-anthropocentric value given to animals' to higher levels of value defended by animal advocates – at the limit, equating the value of an animal life year with that of a human life year. Further, we will propose a sensitivity test that can easily couple the new valuation method we develop below with the existing valuation methods in standard DAMs, which are generally in terms of either dollars or (less frequently) human wellbeing (where human wellbeing estimates involve either a multi-criteria estimate of human health and development, or more typically involve an estimate of wellbeing as a concave function of income or other dollar metric).

The upshot will be a general method for testing the sensitivity of decision analysis to the addition of non-anthropocentric valuation of animals, in a format that can easily be added easily by researchers 'on top of' existing anthropocentric valuation in terms of dollars or wellbeing.

3. Isolating the main problem to be solved: the need for empirically-based interspecies comparisons on a single wellbeing scale

At first glance, it might seem that quantifying animal welfare on the same scale as human welfare poses difficult theoretical challenges. However, there is a sense in which, from an economic and philosophical perspective, there is arguably little theoretical challenge – especially for analysists who already accept the pragmatic use of tradeoff-making aggregations with interpersonal comparisons. For example, from the total utilitarian perspective that is widespread in current analyses, the core theoretical question of how animal wellbeing should be taken into proper account is easy: namely, that Anthropocentric Total Utilitarianism (assumed by most current models):

(1)

$$W^{TU(Anthropocentric)} = \sum_{i \in humans} u_i$$

should be replaced by Interspecies Total Utilitarianism, which takes animal wellbeing into account in addition to human wellbeing:

(2)

$$W^{TU(Interspecies)} = \sum_{s \in species} \sum_{i \in s} u_{is}$$

where i and s are individual and species indices respectively.

In contrast to theoretical questions such as these that are relatively easy, the pressing question that needs to be answered turns out to be a more practical question: namely, how do we estimate e.g. the utility level of a bird, vs. the utility level of a human? This is the most difficult problem that needs to be solved, and theoretical formalism such as that above does nothing to help us solve it.

It is important to see that the problem is *not* how to make *inter-individual* comparisons (interpersonal comparisons), as we already have well-accepted practical methods for making interpersonal comparisons, based on proxies for human wellbeing (e.g. based on consumption levels, which are assumed to be a good estimate of individual utility levels via a uniform diminishing marginal utility of consumption).

Instead, the difficult problem is that we can't yet make even approximate *interspecies* comparisons with confidence, because we don't have well-accepted practical methods for estimating eg the average wellbeing for a bird ($\bar{u}_{bird}$) relative to an average human ($\bar{u}_{human}$).

In the next section we introduce our proposal for solving this central problem of how to make interspecies comparisons.

4. Method: Totalism with Interspecies Comparison Proxy and a Sensitivity Test

In economics, the challenge of making interpersonal comparisons is familiar. Within the discipline, it is also familiar how this problem is solved in practice: namely, by making simply making comparisons based on a method that is believed to involve a good approximation, i.e. by taking consumption (or income, or wealth) as an imperfect but adequate proxy for utility.

Our method for solving the problem of intertheoretic comparisons is analogous to this familiar method from economics. To review, the familiar approach in economics for making interpersonal comparisons is to settle on a single concave function $u(\cdot)$ mapping consumption c to utility, such as:

(3)

$$W^{TU} \approx \sum_{i \in humans} u(c_i) = \sum_{i \in humans} \frac{(c_i)^{1-\theta}}{1-\theta}$$

where θ parameterizes the diminishing marginal utility of consumption.

Often, economic practice is even more crude and approximate than this, because economists often use population-level average consumption $\bar{c}$ as the proxy for the consumption of every individual despite known inequality in individual consumption, as in the following equation that simply multiplies the utility of per capita average consumption by the size of the human population P_h:

(4)

$$W^{TU} \approx P_h \frac{(\bar{c})^{1-\theta}}{1-\theta}$$

The concavity of the utility transformation tells us that this estimate in equation 4 will be inaccurate insofar as inequality exists, and that the inaccuracy of equation 4 versus equation 3 will be problematic insofar as inequality is large, poorly measured and compensated for, or both (i.e. Jensen's inequality holds, as long as $\theta \neq 0$, which would deny diminishing marginal utility of consumption). Economists who use equation 4 to evaluate policies understand that it is theoretically imperfect. So, good judgment is needed as to whether an approximation as in equation 4 is good enough in a particular context, or whether instead we must strive to capture heterogeneity in the way made possible by equation 3.

Our proposal is analogous to the methods used in equations 3 and 4: we propose to make interspecies comparisons based on a proxy that is imperfect but yet is as good as is currently possible. To do this we first need a proxy, call it n, to use as the basis for estimating utility potentials across species. One empirically attractive idea in some contexts is to set n equal to the number of neurons in the brain of members of a species. Data on number of neurons is readily available, and is a good proxy in the context of an enormous global analysis involving billions of individuals where different species are crudely lumped together in small number of bins such as ‘mammals’ and ‘insects’. Alternatively, when greater accuracy is required for specific species or individuals, researchers might set n equal to a more complex metric based on expert analysis of the physical properties of the brains that are best correlated with different levels of cognitive capacity and hedonic enjoyment – e.g. the number of neocortex-like neurons, cortisol levels, or other leading factors identified by the scientific community as most closely correlated with the scalar property of capacity to have complex thoughts and feelings, and whatever other empirical properties are found to be necessary for experiencing utility (Olkowicz et al. 2016, Barron and Klein 2016, Dawkins 2012, Appleby et al. 2011, Fraser 2008).

Abstracting for now from those details, the first step of our proposal is to parameterize an empirical proxy n with an exponential weight ψ into comparative utility potentials for different species, on analogy to θ parameterizing consumption into utility in equations 3 and 4 above. The second step is to multiply this estimate of utility potential by a purely descriptive measure of the degree to which this utility potential is actually realized, f (for “fraction of potential”), to yield the desired utility approximations:

(5)

$$W^{TU} \approx \sum_{is} n_{is}^{\psi} f_{is}$$

In practice, it is often more convenient to use species-level averages as the proxy for utility potential, such as average neuron count $\bar{n}$ for species members, which can then be multiplied by the species population P_s:

(6)

$$W^{TU} \approx \sum_{s} P_s \, \bar{n}_s^{\psi} \bar{f}_s$$

Equations 5 and 6 summarize the basis for our proposed method for making interspecies comparisons. They require an empirical proxy for n (e.g. number of neurons, or a more complex empirically-based metric), values for ψ grounded in normative and empirical considerations (on analogy with how values for θ in equations 3 and 4 are grounded in normative and empirical considerations), and empirically-determined values for f (based on empirical facts about how well members of the species are actually doing relative to their potential).

Finally, the last component of our method, foreshadowed above, is to define a standard sensitivity test that, together with the equations above, captures the range of empirically-grounded and principled estimates that represent normative uncertainty over how to estimate the wellbeing of animals of different species. Figure 1 summarizes the standard sensitivity test of this type that we propose for incorporation into global DAMs based on different principled ways of using the parameter ψ to estimate potential utility of a species s as a function of the average number of neurons n in a member of that species:

		Alternative Estimates of Wellbeing Capacity				
Animal	n	Est. 1	Est. 2	Est. 3	Est. 4	Est. 5
Humans	86,000	1	1	1	1	1
Mammals	250	0.002907	0.000008450514	0.002907	0.029	0.000008450514
Birds	150	0.001744	0.000003042185	0.001744	0.017	0.000003042185
Reptile/Ampl	15	0.000174	0.000000030422	0.000174	0.002	0.000000030422
Fish etc	8	0.000093	0.000000008653	0.000093	0	0.000000008653
Insects etc	0.1	0.000001	0.000000000001	0	0	0
	number of neurons in millions	(ψ = 1) (Higher)	(ψ = 2) (Lower)	(ψ = 1) & insects zero value	(10*ψ = 1) & insects and fish zero value	(ψ = 2) & insects zero value

Figure 1. Five alternative estimates of the wellbeing potential of animal life years of different species based on the number of neurons in an average member of the species. Each estimate is expressed in terms of the wellbeing capacity of one human life year, and thus each estimate divides by the estimated wellbeing capacity of one human life year, $\bar{n}_h^{\psi}$. Estimate 1 = $\frac{\bar{n}_s^{\psi}}{\bar{n}_h^{\psi}}$ with ψ set equal to 1 (a higher estimate of the capacity of animals), whereas estimate 2 = $\frac{\bar{n}_s^{\psi}}{\bar{n}_h^{\psi}}$ with ψ set equal to 2 (a lower estimate of the capacity of animals). Estimate 3 and 5 both stipulate that insects have zero capacity for utility (with the rationale that they fall below some critical threshold), but otherwise use estimates 1 and 2 respectively. Estimate 4 assumes both insects and fish have zero capacity but adds a much higher estimate of the capacity of other animals by multiplying the estimate 1 fraction by 10 for mammals, birds, reptiles, and amphibians.

Each estimate puts human life years (which can be estimated via familiar proxies such as equations 3 or 4 above) on the same scale as the life years of animals of different species, and does so in a principled way that is empirically grounded. These alternative estimates represent the range of empirically-grounded and principled views over the wellbeing of animals of different species. It is not desirable to choose between these estimates, as the goal is to test the sensitivity of optimal decisions to this range of different reasonable (and empirically and theoretically principled) estimates.

Note that equation 6 provides a practical method of estimating or approximating the value of the following more theoretically obvious equation that multiplies the population P_s of each species by the average wellbeing of members of that species:

(7)

$$W^{TU} = \sum_s P_s \, \bar{u}_s$$

We cannot directly use equation 7 prior to a method of making interspecies comparisons such as that developed above, as using equation 7 directly would require knowing the value of $\bar{u}_s$ for each species, which would require knowing the answer to the question of how to make interspecies comparisons. Instead, we must first pioneer a pragmatic method for making those comparisons, such as provided by equation 6: namely, to take $\bar{u}_s \approx \bar{n}_s^{\psi} \bar{f}_s$. The concavity/Jensen's inequality problem that arises from economists' standard reliance on population averages in equation 4 and equation 7 are not as worrisome in an interspecies context in equation 6, because there is less within-species inequality in (e.g.) neuron count and utility than there is within-humanity inequality in consumption and utility.

Additional commentary might be useful on the interpretation of the term f in equations 5 and 6. The initial intuition is that while n^{ψ}represents the capacity for utility, f represents the degree to which this capacity for utility is actually realized, and is thus akin to a quality of a life year adjustment. A more specific interpretation might hold that, for an individual i, f_{is} represents (between 1 and 0) the actual flourishing-of-i-for-an-s where 0 is a species-specific critical level, and 1 is full natural flourishing-for-an-s. E.g. $f_{i,bird}$ = 1 would mean having 100% of the utility a bird naturally can have, while $f_{i,bird}$ = 0.5 means having 50% of natural bird utility.

In many applications, we can make a simplifying assumption without much expected loss of accuracy, and set f = 1 for all non-human species. This is because we can assume that population size of a species s will change orders of magnitude more than f_s. Malthusian dynamics will limit variance in f for non-human species (unlike in humans who have complex technology). In other applications, it will be more desirable for a zoological expert in a particular species to estimate f_s impacts, which does not require making interspecies comparisons (ie does not require quantifying utility impacts) (Wong 2016). Other interpretations in other contexts might also be pragmatically useful.

In sum, the method developed in this section allows interspecies comparisons via equations 5 and 6 based on empirically-available estimates of species population dynamics, actual level of flourishing for member of that species, and empirical proxies for wellbeing capacity. Our specific proposal for integration in large scale global DAMs is to use the sensitivity test displayed in Figure 1, which allows evaluation of tradeoffs between animal life years and human life years. In the next section, we provide an example application within a global DAM.

5. Application: Adding intrinsic valuation of animals to the leading model of optimal climate policy

Climate change will have significant impacts on billions of future animals. However, the DAMs that currently guide climate policy are anthropocentric, with no intrinsic value given to animal wellbeing. Thus, climate models provide a particularly useful example of how animal wellbeing could be added to leading DAMs, and how this might affect estimates of optimal policy. In what follows, we implement the method developed in the previous section – totalism with an interspecies comparison proxy and a standard sensitivity text – in the context of a leading climate change DAM, and calculate the consequences for optimal climate policy. We use the DICE model, which is one of the three models used by the USA and other governments to estimate the social cost of carbon (USIAWGSCC), and is the most widely used of these three. It is the same kind of model used in the Stern Review and in debates about the implications of discount rates (Nordhaus 2007, Stern 2006). DICE's default parameter values for time preference and inequality aversion are 1.5% and 1.5, respectively, which we maintain in the application that follows.

Briefly, the DICE model can be used to analyze the optimal tradeoff between investing in climate mitigation, which incurs a cost relatively soon, and permitting climate damages, which incur costs in the more distant future. DICE is a global model that includes an economic component and a geo-physical component that are linked, where the world is modeled as a single aggregate based on median estimates of climate and economic forces. Exogenous economic projections of population (labor) and technology (TFP) generate the world’s gross output (i.e. what output would be if there were no climate damages) via a Cobb-Douglas production function. Pre-mitigation carbon emissions are a function of gross output and an exogenously determined carbon intensity pathway. These carbon emissions can be reduced (“mitigated,” in the climate policy literature) at a cost to gross output through control policies that are selected via a global harmonized carbon price. Any remaining (post-mitigation) carbon emissions are incorporated into the climate module where they influence global temperature and, ultimately, the future economy through climate-related damages. Damages increase quadratically with a change in the global surface temperature and, like mitigation costs, are incurred directly as the loss of a proportion of gross output. Gross output minus the loss of mitigation costs and climate damages is net output (i.e. output actually realized in the model), which influences capital in subsequent periods via a savings rate.

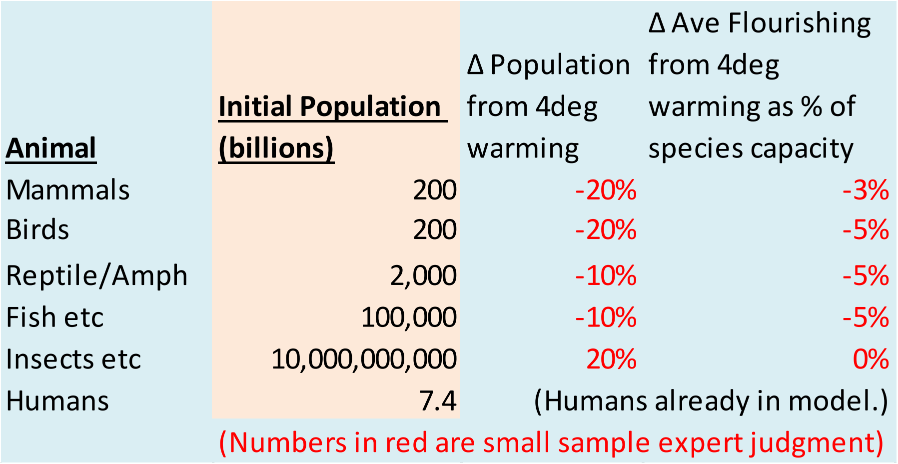
Given this setup, the model's optimization balances mitigation costs, which lower consumption at the time of mitigation, against climate damages which lower consumption in the future. The optimal tradeoff aims to maximize the sum of discounted wellbeing W, which is estimated by the anthropocentric intertemporal discounted utilitarian analog of equation 4 above:

(8)

$$W = \sum_{t} \frac{P_{ht}}{(1+\rho)^{t}} \frac{(\bar{c})^{1-\theta}}{1-\theta}$$

where P_h denotes human population, $\bar{c}$ per capita average consumption, ρ the rate of pure time preference (1.5%), and θ inequality aversion (1.5), and t is a time index. The model is solved by setting carbon prices that maximize this equation, which results in an estimate of optimal emissions reductions via an optimal climate policy that consists of a global carbon price pathway. (For more details of the DICE model, see Nordhaus and Sztorc 2013.)

A precondition for adding intrinsic valuation of animals to a DAM is that impacts on animals must be represented in the model. More specifically, assessment of two kinds of impacts is needed: impacts on (a) populations of animals, and on the (b) wellbeing of those animals. Unfortunately, these impacts are not already represented in leading climate models such as DICE. Fortunately, in the case of climate change and many other examples where ecosystem impacts are not currently estimated in models, it is straightforward to add (based on existing peer reviewed research, or at least expert estimation) at least a reduced-form approximate modeling of these impacts. In the case of climate change, informal conversation with ecologists at leading universities suggests that the following are reasonable central estimates of these two types of impacts:

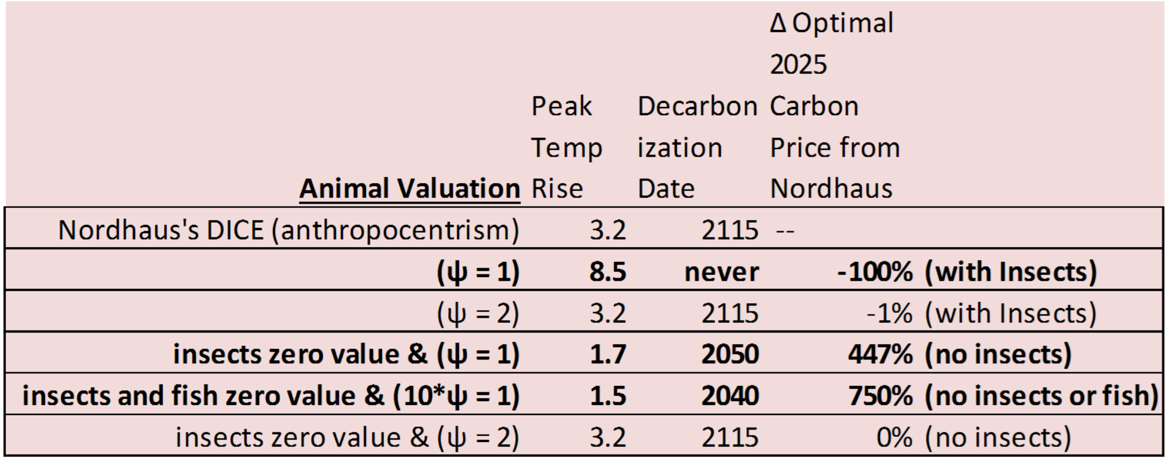


Animal	**Initial Population (billions)**	Δ Population from 4deg warming	Δ Ave Flourishing from 4deg warming as % of species capacity
Mammals	200	-20%	-3%
Birds	200	-20%	-5%
Reptile/Amph	2,000	-10%	-5%
Fish etc	100,000	-10%	-5%
Insects etc	10,000,000,000	20%	0%
Humans	7.4	(Humans already in model.)	
	(Numbers in red are small sample expert judgment)		

Figure 2: Expert estimation of impacts of climate change on animals.

We implement these estimates in a transparent way by simply defining a simple linear function of percent population change and percent flourishing change for each species as a function of temperature increase above preindustrial levels, drawing the line from zero through these estimates of the 4 degree Celsius temperature rise case. This generates the impacts to animals, which are then evaluated using the each of the estimates in Figure 1, which allows simple addition to the total wellbeing calculated via equation 8, resulting in a modified model that has equation 6 as its new objective. (In our modeling, the human utility level that the species utility functions refer to is 2010 average human wellbeing (i.e. animal life years at all time periods are valued as a fraction of the 'baseline' average level of wellbeing for a human in 2010.). See Appendix 5 for further discussion of questions related to this.)

As this illustrates, these methods and the sensitivity test presented above can be used in DAMs to test whether the same direction of effort is required regardless of normative framework, or whether there is a 'sign change' in policy given one normative framework versus another, or whether instead the upshot is more ambiguous (as is illustrated in our results that follow). Here are the results of implementing this sensitivity test in DICE2013:



Animal Valuation	Peak Temp Rise	Decarbonization Date	Δ Optimal 2025 Carbon Price from Nordhaus	
Nordhaus's DICE (anthropocentrism)	3.2	2115	--	
(ψ = 1)	**8.5**	**never**	**-100%**	**(with Insects)**
(ψ = 2)	3.2	2115	-1%	(with Insects)
insects zero value & (ψ = 1)	**1.7**	**2050**	**447%**	**(no insects)**
insects and fish zero value & (10*ψ = 1)	**1.5**	**2040**	**750%**	**(no insects or fish)**
insects zero value & (ψ = 2)	3.2	2115	0%	(no insects)

Figure 3: Sensitivity test of optimal climate change policy in DICE2013 to different assignments of value to the wellbeing of non-human animals.

As these results indicate, climate policy is sensitive to the addition of non-anthropecentric value of animal wellbeing, but in a way that depends highly on which valuation is chosen, and especially depending on how much value is assigned to insects.

One might think a priori that assigning any positive value to the lives of insects would dominate the analysis because there are many quadrillions of insects, and they are estimated to benefit in numbers and wellbeing from climate change (as Sebo forthcoming argues). However, a wide range of optimal policy results are possible even with insects fully incorporated. Interestingly, the results here show that this is not true for the principled valuation scheme ψ = 2, which assigns value to insects using the same function of number of neurons as it uses to assign value to other animals, and in a way that seems to capture arguably a very common view in society about how to assign comparative weight to the interests of animals vs. humans. Still, under ψ = 1 (a very high valuation of animals that is rarely encountered in society), it is true that insects can dominate the calculation because of their sheer numbers together with that higher valuation of their lives. In sum, adding valuation of animal wellbeing in a principled way does not necessarily lead to a 'repugnant conclusion' if insects are given positive utility in a principled way (ψ = 2) (contrary to the conjecture of Sebo forthcoming and a worry of Singer 2016), but insects can indeed 'repugnantly' dominate at higher valuations of 'lower' animals (eg ψ = 1). This is just one more example of the possibility for *every approach* to tradeoff-making social evaluation to yield repugnant-seeming outcomes, which applied to large numbers (Budolfson and Spears forthcoming).

Stepping back from those specifics, perhaps the main upshot is that optimal policy can be wildly sensitive to how value is assigned to animal wellbeing. In this experiment, climate policy is much more sensitive to this factor within the range of values widely defended in society than it is to any other known factor, including discount rates (Nordhaus 2007), human population valuation (Scovronick et al. 2017), and the representation of socioeconomic inequality in climate impacts (Dennig et al. 2015, Budolfson et al. 2018).

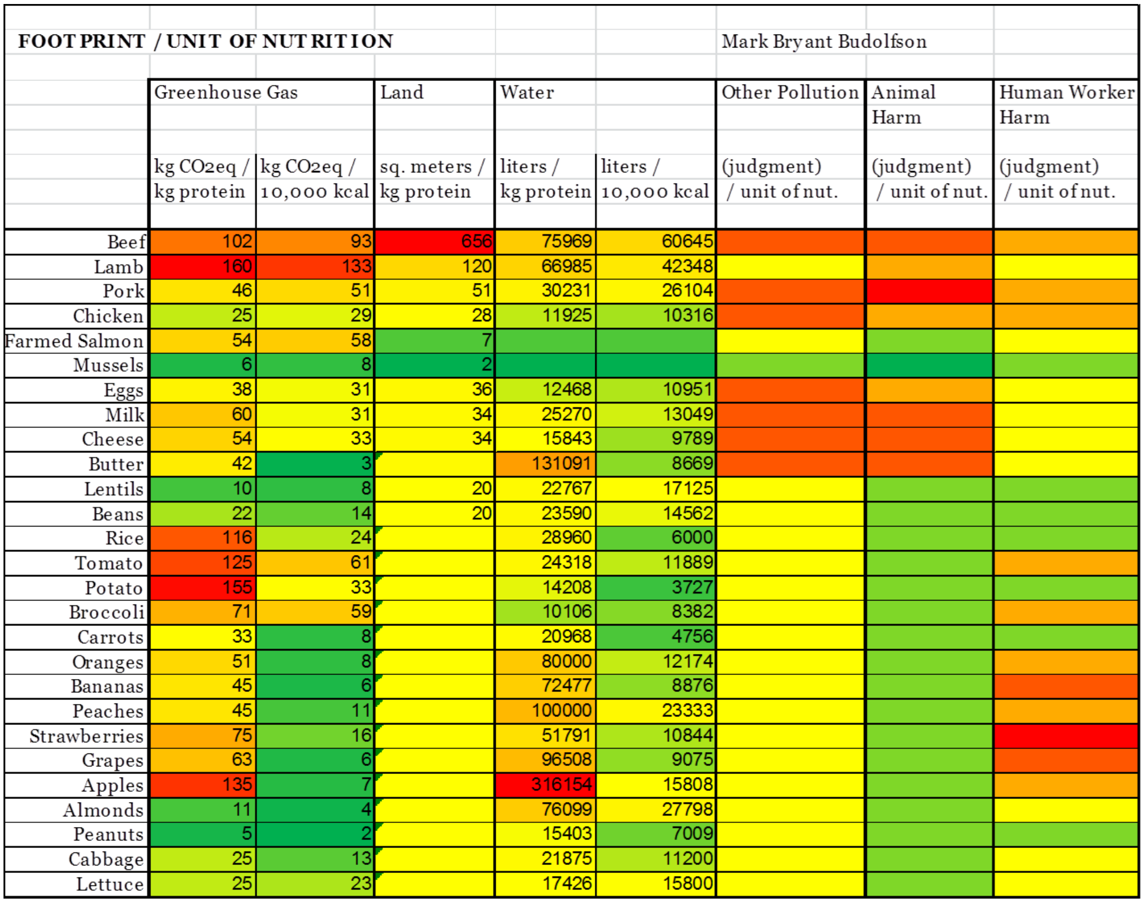
6. Conclusion: Improving valuation of animal wellbeing in policy and decision models

In this paper we've focused on a articulating a pragmatic method and a clear proposal for improving valuation of animal wellbeing in DAMs. The method is motivated by straightforward and established methods in economics for comparing wellbeing between individuals, and for representing alternative but contested normative frameworks within DAMs. The main methodological contribution of the paper has been to develop this method to an extent that it can actually be readily implemented in DAMs used to evaluate policies for society, and that are used in other domains, such as to evaluate decisions such as investments, e.g. in the budding 'effective altruism' movement (GiveWell).

As an application of this method, we've introduced non-anthropocentric valuation of animal wellbeing in climate DAMs, and calculated the implications for optimal policy. The results suggest that optimal policy can be wildly sensitive to how value is assigned to animal wellbeing, and that taking animal wellbeing into account can be at least as important as other known factors.

Appendix 1: More general applications: global sustainability challenges (e.g. the food/water/climate/energy nexus) and the need for assessing tradeoffs between different dimensions of harm to humans, non-human animals, and ecosystems

It is increasingly common for sustainability challenges to be evaluated with global assessment model DAMs. In addition, it is at least as difficult to assess tradeoffs between competing goals in these models as in the case of climate change models, because it is at least as difficult to put the competing values at stake on a common scale. For example, the following chart summarizing some of the tradeoffs involved in selecting a portfolio of food to produce to feed society is indicative of these tradeoffs between human wellbeing, non-human animal wellbeing, and ecosystem health:



FOOTPRINT / UNIT OF NUTRITION						Mark Bryant Budolfson		
	Greenhouse Gas		Land	Water		Other Pollution	Animal Harm	Human Worker Harm
	kg CO2eq / kg protein	kg CO2eq / 10,000 kcal	sq. meters / kg protein	liters / kg protein	liters / 10,000 kcal	(judgment) / unit of nut.	(judgment) / unit of nut.	(judgment) / unit of nut.
Beef	102	93	656	75969	60645	🟥	🟥	🟧
Lamb	160	133	120	66985	42348	🟨	🟧	🟨
Pork	46	51	51	30231	26104	🟥	🟥	🟧
Chicken	25	29	28	11925	10316	🟥	🟧	🟧
Farmed Salmon	54	58	7			🟨	🟩	🟨
Mussels	6	8	2			🟩	🟩	🟩
Eggs	38	31	36	12468	10951	🟥	🟧	🟨
Milk	60	31	34	25270	13049	🟥	🟥	🟨
Cheese	54	33	34	15843	9789	🟥	🟥	🟨
Butter	42	3		131091	8669	🟥	🟥	🟨
Lentils	10	8	20	22767	17125	🟨	🟩	🟩
Beans	22	14	20	23590	14562	🟨	🟩	🟩
Rice	116	24		28960	6000	🟨	🟩	🟩
Tomato	125	61		24318	11889	🟨	🟩	🟧
Potato	155	33		14208	3727	🟨	🟩	🟩
Broccoli	71	59		10106	8382	🟨	🟩	🟧
Carrots	33	8		20968	4756	🟨	🟩	🟩
Oranges	51	8		80000	12174	🟨	🟩	🟧
Bananas	45	6		72477	8876	🟨	🟩	🟥
Peaches	45	11		100000	23333	🟨	🟩	🟧
Strawberries	75	16		51791	10844	🟨	🟩	🟥
Grapes	63	6		96508	9075	🟨	🟩	🟥
Apples	135	7		316154	15808	🟨	🟩	🟧
Almonds	11	4		76099	27798	🟨	🟩	🟨
Peanuts	5	2		15403	7009	🟨	🟩	🟩
Cabbage	25	13		21875	11200	🟨	🟩	🟨
Lettuce	25	23		17426	15800	🟨	🟩	🟨

Figure 4: Average harm footprint among different dimensions of different foods. Numbers based on global averages where available, or USA averages where global

not available. Cells without numbers are based on judgment by one author (from Budolfson 2015).

Clearly, what is needed to move beyond summary statistics such as these to policy recommendations is a DAM that includes, crucially, valuation of these different competing values, and thus provides clear guidance as to how tradeoffs should be made between them, as each of the valuations does in the sensitivity tests above. In this way, the approach above is essential as well to creating adequate global assessment DAMs for evaluating the food/water/climate/energy nexus.

Note also that other areas of debate such as consumer ethics also require a similar analysis of how to make tradeoffs in decisions about, eg. what to eat when these competing values are at stake (Budolfson 2015).

Appendix 2: Total vs. average utilitarianism, and why average utilitarianism doesn't make sense in an interspecies context

Total Utilitarianism (TU) is assumed as the objective to be maximized in standard climate models and in the modeling above. The repugnant-like conclusion therefore threatens with e.g. insects, as investigated above. A related implication of TU noted by Torbjorn Tannsjo (Tannsjo 2016) is that TU seems to imply that a future with intensive factory farming of animals is better than a future with very little animal agriculture, even if it leads to very unpleasant lives for animals, as long as those lives are worth living.

In light of this, many commentators, including effective altruists, have claimed that maximizing total wellbeing is thus implausible as an objective for society when animal wellbeing is taken into account. Instead, they often suggest that Average Utilitarianism (AU) is a more plausible overall objective within which animal wellbeing should be incorporated (GiveWell).

However, we believe this response is incoherent. The problem is that it is not clear how even to define average utilitarianism in a multiple species context, because it is not clear what the denominator is supposed to be. Eg adding one cute beetle with marginally higher-than-average-for-an-insect utility could: (a) improve average wellbeing of every species, (b) improve wellbeing for every individual and create a new individual with a life worth living, but (c) decrease average wellbeing if (as on the most natural proposal) the numerator is total utility and the denominator is all individuals of all species. However, it seems intuitively perverse to say this makes the outcome worse assuming both (a) and (b) obtain.

The underlying issue is that there is no good answer to the question, what is the denominator? To see why, consider two possibilities: either the denominator is unweighted by species (in a sense we will make clear in a moment), or it is weighted; either way, absurdity results. First, consider the unweighted version of interspecies AU:

(9)

$$W^{AU(unweighted)} = \frac{\sum_s P_s \bar{u}_s}{\sum_s P_s}$$

AU(unweighted) fails for the reasons already given, namely that improving the average wellbeing of every species, while making no individual worse off, by adding some flourishing individuals, could make things worse (according to $W^{AU(unweighted)}$).

Now consider the version of interspecies AU that is weighted by species using the sort of proxies for wellbeing capacity introduced in equations 5 and 6 above:

(10)

$$W^{AU(species\ weights)} = \frac{\sum_s P_s \bar{n}_s^{\psi} \bar{u}_s}{\sum_s P_s \bar{n}_s^{\psi}}$$

Here is may be useful to note, as motivation, that adding a species weighting via n_s^{ψ} in the denominator functining as a species weight might seem to help, as it solves one problem: namely, that increases in the size of low-neuron species must bring the overall average down, even if the average of each species is at least as good.

However, this also fails because it implies that improving the average wellbeing of every species, while making no individual worse off, and creating a new individual better off than any existing individual, could make things worse (according to $W^{AU(species\ weights)}$). E.g. adding a high-neuron chimp that is only very slightly better off than its nearly-fully-flourishing chimp compatriots, while leaving the wellbeing of all of many more, flourishing low-neuron animals unchanged. This implication is unacceptable.

So, either way, with or without species weights, averagism fails because it can't give a plausible answer to the question, what's the denominator"?

The upshot is that averagism has a decisive *theoretical* problem associated with its denominator in a multi-species context – and totalism does not have this problem because totalism does not have a denominator. So, unlike other population ethics, averagism cannot plausibly handle interspecies considerations.

Here are some more problems for AU: we don't even know how to define average utilitarianism in an intertemporal context, setting aside interspecies context (Dasgupta 2001). Not to mention all of the other more basic problems for average utilitarianism from Parfit and others (Arrhenius forthcoming, Parfit 1984). And there are new arguments that all non-TU population ethics will also have the repugnant conclusion, or at least have equally repugnant implications (Budolfson and Spears under review).

So, in addition to AU being incoherent in connection with animals (despite the attraction of effective altruists and others), AU also seems to have no advantage over other population ethics including TU in any other way.

Appendix 3: Wong's equation and interspecies comparisons

Effective altruism has an unsolved problem: it needs to evaluate the comparative utility of investments in human-focused vs. non-human animal-focused charities. E.g. how to value investments in animal welfare charities vs. investments in human welfare charities e.g. bednets, anti-worming pills, etc. Kevin Wong in important recent work (Wong 2016) argues that a fruitful assumption is that the structure of the right answer is:

(11)

Δ Quantity of wellbeing = Δ Wellbeing as % of species capacity * Δ life years * Species wellbeing capacity

Wong notes the difficult question is what the value of species wellbeing capacity is for different animals, and he has nothing to say about what it actually is. He calculates the valuation that is implicit in some effective altruism circles, without taking a stand on what the correct evaluation is. For this reason, the discussion above goes beyond the frontier established by Wong's work. At the same time, Wong clarifies the issues and makes a number of important contributions beyond noting the appeal in practice of the equation above. (Wong does not endorse this equation as theoretically correct, only a useful approximation – much like our analysis above.)

Wong's equation is consistent with the approach we used above. The sensitivity test we proposed captures a range of answers to the difficult question of what the green parameter should be. E.g. the 'number of neurons method' indicates how one might ground a practical theory. Thus, consistent with the analyses of Wong and others, effective altruism evaluations can use the same kind of sensitivity test analysis.

Appendix 4: What is the correct utility function to use with animals?

One key issue here: should insects be assigned wellbeing? (Sebo forthcoming, Tye 2017, Barron and Klein 2016, Mendl and Paul 2016, Singer 2016).

We don't think the answer to this question is obvious. Here are some possible arguments.

An argument they should not (we don't necessarily endorse): Philosophical premise: A necessary condition for consideration is consciousness, or some coherent self-conception and preference for and against being in valenced states, or at least subjective experience of those states in some more basic sense, and not merely being in states that are 'valenced' in the most minimal 'proto' sense. Empirical premise: insects are not capable of the necessary state. Therefore, animals should not be assigned any positive wellbeing. (An example of 'merely being in a valenced state, without having the necessary attitude toward that state' might be the insects as described in Mendl and Paul 2016.)

More general possible view (eg Tooley 1972) that could be used in different species utility function development: eg it matters the degree to which one has conscious and cognitive preference, and ultimately a sophisticated plan (or two type: hedonic vs plan satisfaction?)

There are obviously more related arguments on this topic – the preceding gives some of the main ideas behind those that will occur quickly to many.

But maybe the answer is more simple – especially since we've seen that assigning positive value to insects does not necessarily imply a repugnant conclusion, contrary to what is conjectured by Sebo forthcoming... So, perhaps the answer to the question of what is the correct substantive utility function to use with animals is easy: ordinary intuition supports approximately $\psi = 2$?

Appendix 5: What is the human utility level that the species utility functions should refer to?

We used 2010 average human wellbeing in our modeling (i.e. animal life years at all time periods are valued as a fraction of the 'baseline' average level of wellbeing for a human in 2010.)

The reason why it is not obvious what to do in connection with this is that all the evidence indicates that increasing human wealth is correlated without known limit to an increase in human wellbeing. (Philosophers are often misinformed into thinking this is not the case, but see below.) But wildlife wellbeing doesn't seem to increase like that – ie, there isn't a similar mechanism to increasing wealth that causes the wellbeing of wildlife to steadily increase with no obvious limit in sight. So, what human wellbeing level should the animal 'life year discount factor' refer to in the valuation of non-human animal life years? No obvious answer presents itself, not even in outline. That is the problem here.

Side note on why there is not a plateau to human wellbeing: philosophers often mistakenly think that human wellbeing plateaus at a particular point, in connection with the mistaken idea of the so-called 'Easterlin paradox'. For example, see the first graph from Deaton 2013 below, which is the sort of graph displayed in favor of this mistaken idea. Then see how Deaton puts the very same data on a log scale in the next graph, and there is simply a linear line that appears, with increasing wellbeing as a function of wealth. The third graph shows that this is unchanged when one considers a very different proxy for wellbeing than reported life satisfaction, namely, life expectancy.

18 Introduction

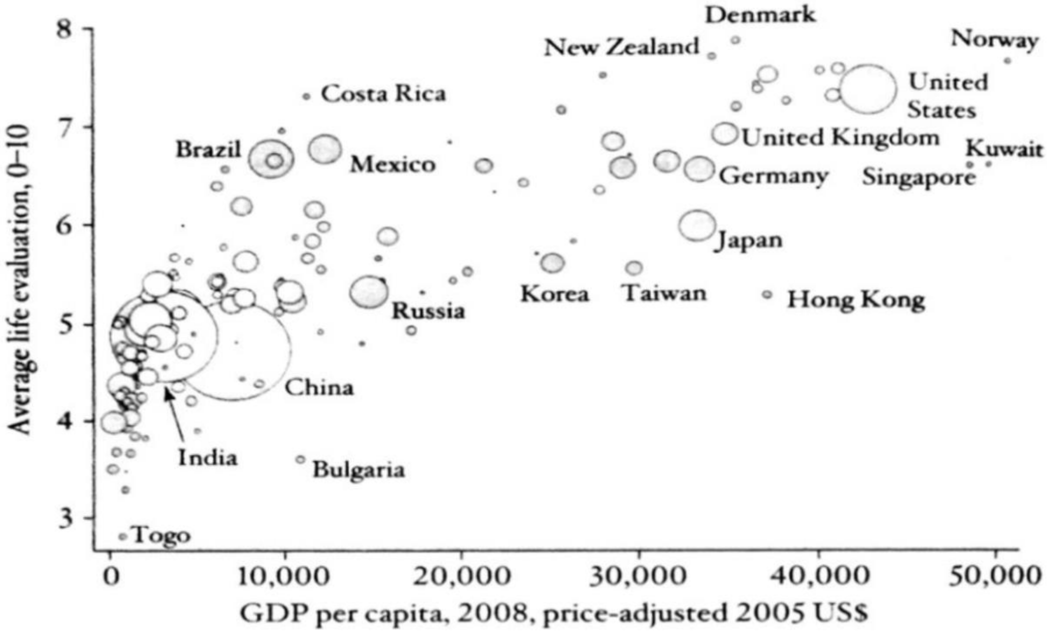


FIGURE 1 Life evaluation and GDP per capita.

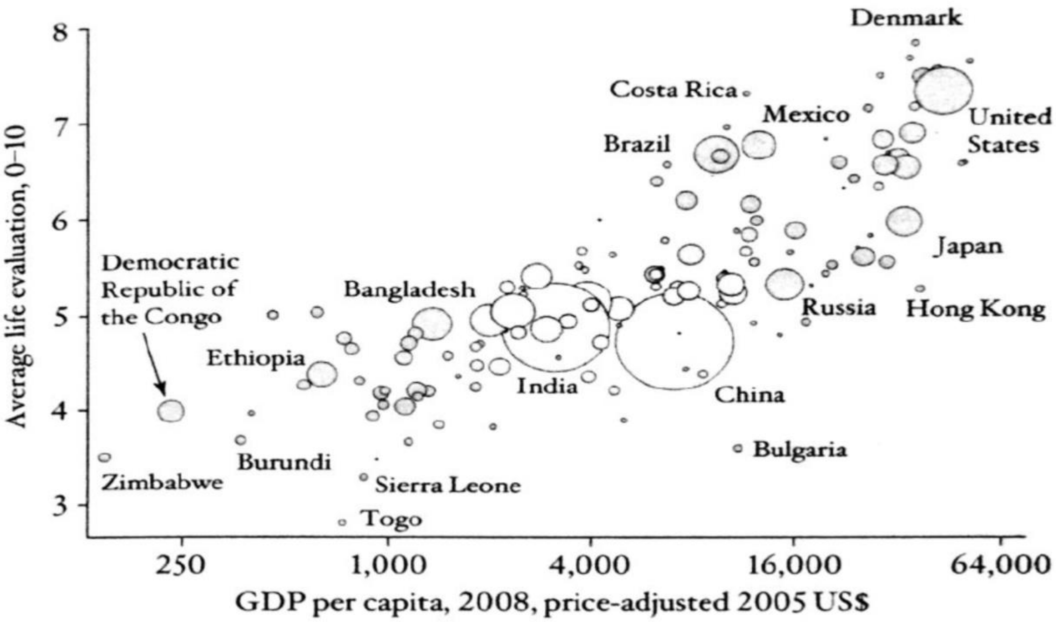


FIGURE 2 Life evaluation and GDP per capita on a log scale.

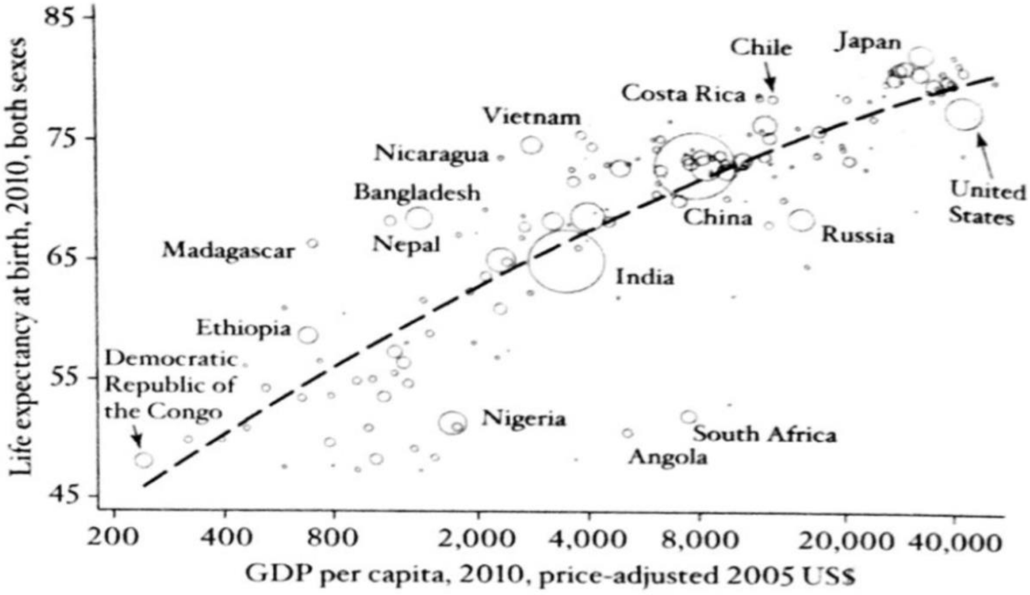


FIGURE 2 Life expectancy and GDP per capita in 2010 on a log scale.

Perhaps more importantly, see Stevensen and Wolfers 2013 and other papers by those authors for more detailed debunking of the so-called ‘Easterlin paradox’.

[END]

References

Alcamo, Joseph. 2003. *Ecosystems and human well-being: a framework for assessment*. Washington DC: Island Press.

Appleby, Michael, Joy Mench, I. Anna Olsson, Francisco Galindo, eds. 2011. *Animal Welfare*, 2nd Edition. CABI.

Arrhenius, Gustaf. forthcoming. *Population Ethics*. Oxford: Oxford UP.

Barron, Andrew and Colin Klein. 2016. "What insects can tell us about the origins of consciousness", *Proceedings of the National Academy of Sciences* 113(18): 4900–4908.

Budolfson, Mark. 2015, "Consumer ethics, harm footprints, and the empirical dimension of food choices", in Chignell, Cuneo, Halteman eds. *Philosophy Comes to Dinner*, Routledge.

Budolfson, Mark and Dean Spears, unpublished. Why all plausible population axiologies imply the repugnant conclusion: a proof.

Budolfson, et al WBER 2018

Chan, Kai et al. 2016. Why protect nature? Rethinking values and the environment. *Proceedings of the National Academy of Sciences* 113(6): 1462-1465.

Costanza, Robert et al. 1997. The value of the world's ecosystem services and natural capital. *Nature* 387: 253-260.

Cowen, Tyler. 2006. Market Failure for the Treatment of Animals. *Society* 43(2):39-44.

Dasgupta, Partha. 2014. Measuring the Wealth of Nations. Annual Review of Resource Economics 6: 17-31.

Dasgupta, Partha on Stern Review

Dasgupta, Partha. 2001. Human Well-Being and the Natural Environment, Oxford UP

Dawkins, Marion. 1980. *Animal Suffering*. Springer.

Dawkins, Marion. 2012. *Why Animals Matter: Animal Consciousness, Animal Welfare, and Human Well-being*. Oxford: Oxford UP.

Deaton, Angus. 2013. *The Great Escape*. Princeton: Princeton UP.

Dennig et al. 2015. Inequality, climate impacts on the future poor, and carbon prices. *Proceedings of the National Academy of Sciences* 112(52): 15827-15832.

Donaldson, Sue and Will Kymlika, 2013. *Zoopolis: A Political Theory of Animal Rights*. Oxford: Oxford UP.

Eyal Frank and Wolfram Schlenker, 2016. "Balancing economic and ecological goals", *Science* 353: 651-652.

Fleurbaey, Marc et al. eds., International Panel on Social Progress (IPSP), 2018. *Rethinking Society for the 21st Century*. Cambridge: Cambridge UP.

Fraser, David. 2008. *Understanding Animal Welfare: The Science in its Cultural Context*. Wiley.

Gillingham, Kenneth, William Nordhaus, David Anthoff, Geoffrey Blanford, Valentina Bosetti, Peter Christensen, Haewon McJeon, and John Reilly. 2018. Modeling Uncertainty in Integrated Assessment of Climate Change: A Multi-Model Comparison. *Journal of the Association of Environmental and Resource Economists*.

GiveWell. www.givewell.org see also good ventures

Greaves, Hilary. 2017. "Population Axiology", *Philosophy Compass* 12:e12442.

Hiller, Avram.

Jamieson, Dale. 2008. *Ethics and the Environment*. Cambridge: Cambridge UP.

Klein, Colin and Andrew Barron. 2016. Insect consciousness: Commitments, conflicts and consequences. *Animal Sentience* 9(21).

Klein, Colin and Andrew Barron. 2016. Reply to Adamo, Key et al., and Schilling and Cruse: Crawling around the hard problem of consciousness. *Proceedings of the National Academy of Sciences* 113 (27) E3814-E3815.

Mendl and Paul (2016) "Bee happy: Bumblebees show decision-making that reflects emotion-like states", *Science* 353: 1499-1500.

McShane, Katie. 2018

Ng, Yew-Kwang. 2016. How welfare biology and commonsense may help to reduce animal suffering. *Animal Sentience*.

Ng, Yew-Kwang. 1995. Towards Welfare Biology: Evolutionary Economics of Animal Consciousness and Suffering. *Biology and Philosophy* 10: 255-285.

Nordhaus review of Stern review 2007

Nordhaus, William and Paul Sztorc. 2013. DICE 2013R: Introduction and user's manual. Online at: http://www.econ.yale.edu/~nordhaus/homepage/homepage/documents/DICE_Manual_100413r1.pdf

Norwood, F.B, and J.L. Lusk. 2011. *Compassion by the Pound: The Economics of Farm Animal Welfare*. Oxford: Oxford UP.

Olkowicz, Seweryn et al. 2016. Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences* 113 (26): 7255-7260.

O'Neill, John, Alan Holland, and Andrew Light. 2008. *Environmental Values*. New York: Routledge.

Palmer, Clare, Katie McShane & Ronald Sandler. 2014. Environmental Ethics, *Annual Review of Environment and Resources* 39:419-442.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford UP

Rawls, John. 1972. *A Theory of Justice*. Cambridge: Harvard UP.

Sandler, Ronald. 2018. *Environmental Ethics: Theory in Practice*. Oxford: Oxford UP.

Sarkar, Sahotra. 2012. *Environmental Philosophy: From Theory to Practice*. Wiley-Blackwell.

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge: Harvard UP.

Schmidtz, David and Dan Shahar eds. 2018. *Environmental Ethics*, 3rd edition. Oxford: Oxford UP.

Scovronick et al. 2017. Impact of population growth and population ethics on climate change mitigation policy, *Proceedings of the National Academy of Sciences* 114(46): 12338-12343.

Sebo, Jeff. forthcoming "Animals and Climate Change", in Budolfson, McPherson, and Plunkett eds. *Climate Change and Philosophy*. Oxford: Oxford UP.

Singer, Peter. 1975. *Animal Liberation*. HarperCollins.

Singer, Peter. 2016. "Are insects conscious?", Project Syndicate. Online at: https://www.project-syndicate.org/commentary/are-insects-conscious-by-peter-singer-2016-05

Shriver, Adam.

Stern Review

Stevenson, Betsey and Justin Wolfers. 2013. Subjective Well-Being and Income: Is There Any Evidence of Satiation? *American Economic Review* 103(3): 598–604.

Sunstein et al climate change and animals

Sunstein, Cass and Martha Nussbaum, eds. 2005. *Animal Rights: Current Debates and New Directions*. Oxford: Oxford UP.

Tannsjo, Torbjorn. 2016. It's Getting Better All the Time, in Barnhill, Budolfson, Doggett eds. *Food, Ethics, and Society*. Oxford: Oxford UP.

Tooley, Michael. 1972. Abortion and Infanticide. *Philosophy & Public Affairs* 2: 37-65.

Tye, Michael. 2017. *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* Oxford: Oxford UP.

Weitzman on Stern

Wong, Kevin. 2016. *Counting Animals: On Effective Altruism and the Prospect of Interspecies Commensurability*. Princeton University, Honors Thesis.

USIAWGSCC, U.S. interagency working group on the social cost of carbon