Public Policy, Consequentialism, the Environment, and Non-Human Animals

Mark Budolfson and Dean Spears

ABSTRACT

The focus of this chapter is public policy and consequentialism, especially issues that arise in connection with the environment – i.e. the natural world, including non-human animals. We integrate some of the existing literature on environmental economics, welfare economics, and policy with the literature on environmental values and philosophy. The emphasis on environmental policy is motivated by the fact that it is arguably the most philosophically interesting and challenging application of consequentialism to policy, as it includes all the challenges of valuing the distribution of human wealth and power, and has the further challenge of putting these consequences on the same scale as consequences for human health, non-human animals, and nature. We suggest that standard methods of (economic) policy analysis provide a good approximation of correct welfarist analysis, except that they must be supplemented with methods for valuing animal wellbeing and tradeoffs with human wellbeing. We then provide the needed methods.

KEYWORDS

Consequentialism, utilitarianism, policy, regulation, environment, animals, pollution, welfare, wellbeing, axiology

1. CONSEQUENTIALISM AND THE ENVIRONMENT

Consequentialism evaluates alternative possible courses of action ('policies' or 'options' or 'choices', for short), estimates their comparative desirability, and is a leading framework for answering questions about what choices are better and worse. The focus of this chapter is on distinctive issues for consequentialism and public policy, especially those that arise in connection with the environment – i.e. the natural world, and especially non-human animals.

The application of consequentialism to real-world decisions has three main structural components, where the latter two differ in substance between different substantive consequentialist views. These three structural components are:

> (1) **impact assessment** (i.e. how different options are likely to lead to different consequences),
>
> (2) **axiology** (i.e. how to aggregate the consequences within each possible outcome into an aggregate evaluation of the comparative goodness and badness of each outcome), and
>
> (3) **decision theory** (i.e. how to rank options based on the probability of different outcomes conditional on their choice together with the comparative goodness and badness of those outcomes).

As an illustration, these factors are recognizable in familiar consequentialist debates about the ethics of individual environmentalist choices such as whether one must avoid eating meat, whether one must reduce one's carbon footprint, and the like. In each case, consequentialist analysis turns on (1) impacts (e.g. whether it makes any difference to animal suffering whether a single person buys meat or not) (Singer 1980, Norcross 2004, Kagan 2011, Budolfson 2018, Nefsky 2019), (2) how to value outcomes based on impacts (e.g. whether impacts that reduce billions of people's lives by an imperceptible amount of time could add up to something as bad as killing a normal adult in the prime of life) (Nolt 2011, Broome 2012, Budolfson 2012, Kagan 2011, Nefsky 2011), and (3) how these factors should determine what you must do (e.g. whether you are required to choose the option that maximally benefits society in expectation, or whether it is permissible to do much less as long as it is 'good enough') (Singer 1972, Portmore 2011).

Much has been written from a philosophical point of view on consequentialism and these individual-level choices in response to environmental challenges. Partly for this reason, such individual-level choices will not be our focus here. In addition, we set aside individual-level issues here because they are often driven by factors that are not essentially about the environment, but by more general issues about consequentialism and demandingness, or consequentialism and collective action, and so on.

Less has been written from a philosophical point of view about consequentialism and how society should make public policy choices, especially about the environment and non-human animals. This will be our focus in what follows, where we endeavor to integrate some of the existing literature on environmental economics, welfare economics, and policy with the existing literature on environmental values and philosophy.

2. STANDARD POLICY ANALYSIS (SPA)

In applying consequentialism to real-world public policy decisions, the leading method is what might be called *standard policy analysis* (SPA). As an instance of applied consequentialism, SPA combines (1) an impact assessment that models the empirical dynamics that determine outcomes as a function of policy choices (which generates an assignment of probabilities to outcomes conditional on policy choices), (2) an axiology, where SPA assigns value to outcomes based on how the impacts within each outcome are valued by humans, and (3) a decision theory that evaluates the choiceworthiness of policy options as a function of (1) and (2), where SPA assumes a decision theory that has a familiar expectation or maximize expected value form, in which policy options are valued based on the sum of the possible outcomes of each policy option weighted by their probability conditional on the choice of that policy.[1] In many cases, a more partial analysis is performed using methods of SPA, such as a cost-benefit analysis, which might for example evaluate whether the benefits exceed the costs of a particular kind of pollution control in a number of scenarios that model various levels of stringency of control; although this is only a partial analysis that ranks the comparative choiceworthiness of a small number of policy options, the methods of this kind of cost-benefit analysis are typically those of SPA (Drummond et al. 2005).

SPA in this sense is the most influential methodology that informs public policy (Sunstein 2014, Adler and Fleurbaey 2016). SPA is often (and increasingly) used in other contexts beyond public policy, such as by NGOs, individuals, and foundations deciding what initiatives to fund

---

[1] It is possible to endorse a different decision theory. On normative grounds, this has been advocated for different reasons by Buchak 2013 and Portmore 2011; thus it is important to be explicit that 3 is a further assumption independent of 1 and 2. Because they are extensively discussed elsewhere, we will not focus on these alternatives here. We also set aside the important issue of decision-making under normative uncertainty – for discussion, see MacAskill et al. forthcoming and Budolfson and Spears 2019c.

(GiveWell.org 2019), and in 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. SPA is especially pervasive in evaluations of policies that will have widespread socioeconomic consequences, and in the environmental domain. This is not to say that policies enacted by policymakers generally conform to SPA, as instead other values and political objectives often carry the day (Sunstein 2018). Rather, the point is merely that SPA is the most influential analytical input into actual policy analysis, and the use of SPA is widely judged to be normatively correct by widely cited scholars on public policy, even if actual policymakers rarely conform to its recommendations (Sunstein 2014, Sunstein 2018).

In what follows we discuss the main substantive assumptions of SPA in more detail, highlight some common objections to SPA from philosophers and others, and explain the resources that SPA has for replying to those objections, and their limits. In later sections we discuss the prospects for improving SPA.

2.1 IMPACT ASSESSEMENT IN SPA

As may already be clear, SPA factors into empirical and evaluative parts. The empirical part models the dynamics that determine outcomes as a function of policy choices, and is generally based on work from social, health, and/or natural or other empirical sciences. For example, in the case of income tax policy, the dynamics might be taken from economic studies. In a multidisciplinary context such as air pollution policy, the dynamics might be taken from both atmospheric and public health science (namely the benefit side of the equation, based on the science of population-level impacts of different levels of exposure to air pollutants) and economics (namely the cost side of the equation, based on energy economics models of the cost of different levels of pollutant reductions via different policy instruments). In a maximally large-scale problem like climate change, an ‘integrated assessment model’ might have to be developed by teams of scientists from a wide variety of disciplines to model the coupled complex systems involved, and their associated impacts along a multitude of diverse dimensions for different individuals at different locations in space and time, at each point estimating impacts conditional on policy choices for different sectors that drive human wealth, health, migration, and demography, and the wellbeing and population dynamics of flora and fauna, impacts on ecosystems, and so on. This is what is actually done in the case of climate change, and increasingly other global environmental challenges such as ecosystem preservation and the like (IPCC 2014, IPBES 2019).

We set aside the details of impact assessment given its empirical nature, but before doing so it is worth mentioning some of its limitations. The first is simply empirical uncertainty, which increases as estimates extend into the more distant future. Less obvious but important recurring problems are also a frequent inability to anticipate important negative unintended side effects of policies, and a frequent inability to estimate the capacity for innovation or social coordination to endogenously improve outcomes in response to societal challenges (Deaton 2013 chapter 7, Connelly 2008, Lam 2011, Ostrom 1990).

2.2 AXIOLOGY IN SPA: ANTHROPOCENTRIC VALUATION BASED ON HUMAN PREFERENCES

Given possible outcomes modeled by an impact assessment, SPA assigns value to those outcomes based on estimates of the value of the impacts within each outcome to humans, together with a *social welfare function* that aggregates the value of all of these impacts to different humans into a single aggregate societal value of the overall outcome; together, this is the axiology of SPA. One key feature of SPA is that this axiology is *anthropocentric*, in the sense that it bases its valuation on methods of estimating how the relevant impacts would be valued *by humans*, typically measured in terms of impacts on overall GDP or societal wealth, or on as a function of the willingness to pay by individuals to achieve or avoid them.

In making these anthropocentric assumptions, SPA mirrors the normative assumption of mainstream economics about fundamental value, namely that what fundamentally makes for better or worse outcomes is the extent to which those outcomes are preferred or dispreferred by humans. It is important to see that this is indeed a *normative* assumption (since this assumption is used to conclude that some outcomes are *better* than others and thus *should* be chosen by policymakers), and that this assumption is not 'neutral' (contrary to what many economists claim) since for example it implies that we should ignore the wellbeing of non-human animals in a way akin to how colonialists ignore the wellbeing of indigenous people except insofar as what was good for those people aligned with colonial interests.

More generally and beyond the implications for animals, many normative theorists find the axiology assumed by mainstream economics dubious, on the grounds that it ignores the possibility that preference satisfaction might not be the only determinant of individual wellbeing, and because it ignores other possible determinants of better and worse outcomes, including considerations of justice (Hausman et al. 2018). Later sections will add detail to these critiques, evaluate their merit, and identify alternatives to SPA. For now, the next order of business is to better understand the assumptions of SPA so that its resources to reply to these objections can be made more clear.

2.2.1 VALUATION OF IMPACTS IN SPA

Because environmental challenges like air pollution and climate change can have impacts on almost the entire range of things that might be considered fundamentally valuable, it is useful to provide some framework for enumerating these valuable things, and to examine the resources that SPA has or doesn't have to properly evaluate these impacts. Here is a brief and non-exhaustive list:

- consequences for individual humans along the dimensions of:
    - wealth
    - health
    - happiness
    - freedom
    - cultural and aesthetic values
- consequences for individual non-human:
    - sentient animals (both wild and farmed)
    - non-sentient living things such as crops, trees, and so on
    - non-living things such as mountains
- systemic consequences:
    - distribution of wealth, and other human distributional consequences

- distribution of ecosystems of various compositions, and other non-human distributional consequences

This is not meant to be an exhaustive list, but merely indicative of what impacts might be seen as valuable given a substantive philosophical view – this list may be helpful as one considers whether SPA or some alternative approach to policy analysis can properly value all such things.

A common caricature of SPA is that it ignores the value of everything on the list above except for the first thing (wealth), on the grounds that SPA cares only about things that have monetary value in the marketplace. It is important to see that this is a confused criticism, as SPA (especially in the environmental domain) often aims at valuing all of the things listed above, and does so via principled and well-developed methods of valuation based on human preferences, where the essence of the project is to derive the value of things that do not have monetary values in the marketplace from other measures of human preferences. At the same time, it is also true that some applications of SPA are indeed overly simplistic and do not adequately accomplish this goal. In what follows, we aim to give a fair description of how SPA can be used to value things on the list of things above in a principled way based on human preferences, as well as a clear view of how some applications of SPA may fail to achieve this aspiration.

To begin to see why valuing things that have no monetary value in the marketplace is part of the essence of the project of SPA in the environmental realm, note that one general rationale for policy that is endorsed by proponents of SPA is what might be called the *market failure rationale for policy*, namely, that prices in the marketplace sometimes systematically fail to account for some of the costs and benefits to other humans of the transactions that give rise to those prices, and under these conditions unregulated market transactions will not generally lead to the best outcomes for society, and so regulatory intervention is justified. Most on point for our purposes are cases where *negative externalities* of transactions exist, in the sense that some individuals are harmed by a transaction because they are not party to the transaction and thus the price at which the exchange happens does not reflect the strength of their preferences. For example, in the 1950s air pollution emissions were largely unregulated in the UK and the USA, and as a result, when a factory owner produced goods and sold them in the marketplace and created emissions in the process, the preferences for air quality of all the people who were harmed by those emissions were generally ignored in determining the market prices at which the goods were bought and sold. And if everyone's preferences for air quality had been taken into account (i.e. if the producer had to pay everyone harmed by his or her pollution in the same way consumers had to pay the producer for the goods he or she produced, namely, to the point at which individuals were happy to accept the air pollution in exchange for that payment), then some factories would have had to close and the air would have thus ended up cleaner, and the value of the air quality benefits (measured in the aggregate willingness to pay for them across society) would have been more than enough to compensate for the lost production. In this way, the lack of regulation of air pollution in the 1950s created a situation where free markets led to outcomes that were worse than what they could have been if everyone had to pay the true social cost of their pollution. If prices had instead internalized all of the true costs in the way just described, then many people could have been made better off without anyone being made worse off, a *pareto improvement*, generally taken by economists to be an uncontroversial example of a better outcome.

One of the contributions of welfare economics is the proof that under conditions of perfect competition including no externalities, the end result of free exchanges in society would be an outcome that is *pareto optimal* in the sense that no pareto improvements to that outcome would be possible; at the same time, when negative externalities exist or other features of imperfect competition exist such as monopoly power, then there is a clear reason to expect *market failure* in which free exchange would result in a suboptimal outcome in the sense that a pareto improvement would be possible (Kolstad 2010; conceptually this is based on the *First Fundamental Theorem of Welfare Economics*). In the face of market failure, a rationale emerges to use regulatory policy to improve outcomes. It is important to see that the nature of this market failure rationale is *pro tanto* and may be defeated in many cases, because policy should not always be expected to improve outcomes, as government policies often generate even worse outcomes, given that government is imperfect in predictable ways just as unregulated markets are imperfect in predictable ways (Budolfson 2017).

A conceptually instructive policy instrument that is often considered by SPA as a means to correct market failures involving environmental pollution is a *Pigovian tax*, which is price charged to those who impose pollution (or more precisely, negative externalities on others) based on a calculation of the aggregate cost to society (the *social cost*) of that type of pollution, measured in monetary terms by aggregate willingness to pay across society to avoid that type of pollution. Theoretically, if such a tax were levied at the marginal cost to society of an additional unit of the relevant type of pollution at the point at which that marginal cost is equal the marginal benefit to society of an additional unit of that form of pollution, then the optimal level of that form of pollution would result (Kolstad 2010).

Returning now to the long list of potentially valuable impacts above, the key resource of SPA are methods that aim to value all such impacts on a single scale of willingness to pay that represents aggregate human preferences to achieve or avoid those impacts, even when those impacts do not have a market price. When done well and comprehensively, SPA thus promises to value all valuable impacts in a principled way based on human preferences. Furthermore, according to mainstream economics, this is the only normatively acceptable way of valuing impacts, since policies based on any other valuation scheme would illegitimately force outcomes onto society that do not reflect the preferences of the population, and instead involve dictators imposing their own values on the population (Nordhaus 2007 pg. 691). We consider objections to this view further below. For now, we outline in more detail the methods SPA uses to value the impacts enumerated above in connection with a representative range of leading examples in environmental policy and philosophy:

- environmental pollution
- the treatment of non-human animals
- natural resource use and land management
- environmental justice (discussed in the next subsection)

Environmental pollution has been mentioned already, but the valuation of impacts has not been discussed at the level of detail relevant to understanding best practices applications of SPA. In the case of pollution, applying SPA requires identifying and modeling the relevant impacts of pollution and then valuing them. Some of the impacts typically modeled are mortality (e.g. additional deaths from heart attacks as a function of increased exposure particulates in the air) and morbidity (e.g. additional burden of asthma), impacts on recreation and cultural and aesthetic values (e.g. preference for clear skies for aesthetic reasons, and a preference for cathedrals not to be defaced by acid rain), and pollution has impacts on many other things such as crop yields and the like. SPA aims to monetize each of these impacts based on human willingness to pay, and thus put them on the same scale as the benefits from goods produced by the emissions generating activities. The general aim of policy is then to reduce emissions in a way that reflects all of these preferences and associated willingness to pay, down to level at which market failure no longer generates a suboptimal outcome.

In the case of farm animals, SPA also has the resources to provide a market failure rationale for animal welfare improvements, as many economists argue that humans are sufficiently willing to pay for better animal welfare to more than offset the cost of some animal welfare improvements (Norwood and Lusk 2011, Cowen 2006). A different market failure rationale for animal welfare improvements is that they would more than pay for themselves by reducing the expected harm to human health from diseases, antimicrobial resistance, and the like, where these harms to human health are not reflected in the market prices of animal products; thus, policies that included targeted animal welfare improvements could yield benefits for everyone in expectation (Jarvis and Donoso 2018, Otte and Chilonda 2000). In this way, SPA has the resources to argue for substantial animal welfare improvements – namely, to a higher level of animal welfare that best satisfies the preferences of humans.

This provides a useful segue into an outline of general SPA methods for estimating willingness to pay for *non-market* impacts that do not correspond to market prices. Valuing these impacts is a nontrivial challenge. One dimension of this challenge is estimating the *use value* of elements of nature (human willingness to pay to use animals, trees, minerals within a mountain), and another dimension is estimating the *nonuse value* (willingness to pay for such things to remain unused). Use value includes willingness to pay for direct use (e.g. to buy and eat animals, convert trees into housing materials), as well as indirect use, including *ecosystem services* such as the value of pollinators in human agriculture, the value of aquatic mollusks in cleaning water for human use, the value of wildlife to human recreation, and so on. Nonuse value also includes willingness to pay for the continued mere existence of elements of nature without use by humans (*existence* *value*, which is especially important for preservationist valuation of wildlife, biodiversity, wilderness areas), as well as the *option value* of keeping elements of nature around for potential future human use in ways that will turn out to be valuable, but of which we may be currently ignorant (Arrow and Fisher 1974).

Substantive methods are needed to estimate willingness to pay for these things, as their values are often not readily reflected in market prices.[2] *Contingent valuation methods* are generally based on surveys that elicit self-reported willingness to pay to avoid or bring about particular outcomes. However, there are a number of objections to this method, perhaps the most important of which is the worry that it leads to biased and inflated estimates of willingness to pay, because people do not have to back their answers with real investments (Hsiung and Sunstein 2007). This points the way toward the main alternative method, based on *revealed preference methods* for measuring valuations implicit in actual choices, beyond what can be immediately read off from market prices. Revealed preference studies are especially foundational in the valuation of mortality, as willingness to pay to avoid increased occupational risk of death and the like are often used to estimate the value of excess deaths or life years lost. These estimates of willingness to pay for human health are often the dominant factor in estimates of the social cost of pollution and other cases where environmental quality has a clear impact on human health (Sunstein 2014). These methods can also be used to infer willingness to pay for amenities like open space, parks, and the like when those amenities can be seen as partly determining the price of things

[2] In contrast, when the anthropocentric value of animals are well-reflected in market prices – such as e.g. the price of pollination services – market prices are the preferred method of valuation, at least to the extent that the good is a private good traded in a well-functioning marketplace.

like housing that do have clear market prices; in such cases, inferential methods can be used to extract an estimate of the contribution of the amenity to the comparative prices of e.g. houses.

These methods are the most widely-used methods for estimating the value of non-market impacts (Kolstad 2010).

Policy analysis also often requires methods of estimating willingness to pay for resources in the future, and a calculation of the present value of those future benefits, as ignoring these long-run benefits in an unregulated market could create a tragedy of the commons in which resources are used unsustainably with an eye toward only short-term profits, leading to a worse outcome for society than if they were managed in accord with an analysis that accounted for long-run value (Ostrom 1990, Sandler 2018 pp. 113-124).

2.2.2 SOCIAL WELFARE FUNCTIONS IN SPA

Given the valuation within each possible outcome of impacts to individuals by SPA, the next step is to aggregate those into an overall valuation of each outcome, so that the goodness and badness of the different possible outcomes of policy choices can be compared. There is no single agreed upon formula – *social welfare function (SWF)* – for this aggregation. Instead, there are a number of different SWFs that are sometimes used in SPA, which we will describe and contrast. All these SWFs are motivated by the conceptual idea of *individualist anthropocentric welfarism*, in the sense that they share the basic conceptual idea that the goodness of an overall outcome is a principled function of the wellbeing of the individual humans within that outcome, and that individuals have more wellbeing the more they consume, where this notion of an individual's *consumption* is taken to include all the goods and services, leisure time, health, and everything else that the individual values, as measured by their willingness to pay.

An important distinction between different SWFs used in SPA is whether they merely focus on a societal level economic sum (e.g. GDP plus the net monetized value of all of the non-market impacts described in the previous section), or whether instead they represent differences between individuals, such as different levels of individual consumption, differences in race, gender, age, location, and so on, and accordingly estimate different wellbeing consequences when the same impacts affect different individuals, and then aggregate those heterogeneous wellbeing consequences. It is common to use the former method, although the latter is more precise from a normative point of view. Some advantages of the former method are that it is simple and avoids the need for an assumption about how to make interpersonal comparisons of wellbeing levels. However, if one accepts that there is diminishing marginal utility of consumption in the case of any given individual (as every economist does), one would presumably think that at the population level a similar diminishing marginal utility of consumption would arise – i.e. one would assume that at the population level giving an additional dollar of consumption to the poor would tend to increase wellbeing more than giving that additional dollar to the rich. A widely used SWF in SPA that captures this thought is the following *isoelastic utility function*:

(1)

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

For our purposes, the important feature of the SWF in Equation 1 is that there is diminishing marginal utility of consumption, which allows a policy analysis to model the idea that there is greater wellbeing generated by a dollar worth of increased consumption for the poor vs. the rich.

Empirical studies of reported happiness and income provide some evidence for kind of approach as well (Kahneman and Deaton 2010). In addition, this approach avoids the normatively objectionable implication that impacts to the poor are less important than impacts to the rich simply because the poor have lower willingness to pay, which is an implication of the first approach that simply sums monetized values of impacts.

2.2.3 ENLIGHTENED SPA

With all of this in mind, many advocates of SPA believe that best practices should use a SWF analogous to Equation 1 in which individual wellbeing is estimated as a concave transformation of consumption, which can then be aggregated at the societal level in any number of ways that can include principles of distributive justice, while avoiding the implication that the poor are worth less simply because of their lower willingness to pay for health, life, and other goods (Adler and Fleurbaey 2016). Although Equation 1 is a total utilitarian SWF, alternative SWFs exist and have been advocated as best practices within SPA to represent prioritarian, egalitarian, Rawlsian maxi-min, and other methods of aggregation, as well as the range of non-totalist welfarist population ethics (Budolfson and Spears 2018). Because the choice of SWF involves the choice of a population ethics as well, this can have implications for policy in some cases (Scovronick et al. 2017), but not in a way that need be any more dramatic than the familiar way that the choice of the shape of the transformation between consumption and wellbeing has implications for policy (Budolfson and Spears forthcoming). In addition, although it is commonly believed that totalist SWFs have a special liability to the *repugnant conclusion* (when an axiology implies that an enormous number of barely worth living lives can be better than a smaller number of good lives), recent work has suggested that any SWF that endorses tradeoffs between the wellbeing of individuals will have such implications, and so the repugnant conclusion arguably does not tell for or against any welfarist SWF (Budolfson and Spears forthcoming, Budolfson and Spears 2018). (See the discussion of insect valuation and the repugnant conclusion further below.)

The upshot is that SPA need not ignore considerations of distributive justice and future generations, and indeed has the resources to integrate them into policy analysis – we might call such an approach *Enlightened SPA*.[3] At the same time, it should be emphasized that in actual practice it is more common for SPA to use a normatively inferior SWF that aims merely to maximize the sum of GDP, plus some monetized nonmarket impacts – and outside the environmental realm, it is very common for applications of SPA to focus only on maximizing GDP, without any accounting for nonmarket impacts, let alone the socioeconomic distribution of impacts (Stiglitz et al. 2009).

In connection with environmental justice, distinctive issues arise about the distribution of environmental impacts. In particular, a familiar concern is that racial minorities are exposed to more pollution and other environmental inequalities, such as having fewer recreation opportunities and more waste dumps located near their homes (Shrader-Frechette 2002). In response, some would argue that this may be merely a consequence of economic inequality, as the price of homes will be lower in environmentally undesirable locations, and so it is an inevitable result of free choice by the poor to do as best as they can within the marketplace to live disproportionately in such locations (Banzhaf 2009). However, recent work has provided

[3] Compare the concept of *enlightened anthropocentrism* in Sandler 2018.

some evidence that this explanation may be inadequate, as in the USA minorities are disproportionately exposed to air pollution even after controlling for income differences (Mikati et al. 2018). An Enlightened SPA SWF that displays aversion to inequality in health-related impacts could be used to capture the importance of alleviating such injustice.

A further dimension of complexity is that many environmental policy challenges such as climate change require policy analysis over a very long time horizon, with benefits distant in the future from the cost of investments to achieve those benefits. In such a case, a SWF must incorporate key assumptions about how to calculate the present value of the future costs and benefits for society, often referred to as assumptions about *discounting*. One dimension of discounting is the parameter θ above, which parameterizes the diminishing marginal utility of consumption; another dimension of discounting not represented above is the rate of pure time preference, which determines how much less weight should be given to wellbeing consequences in the future simply because they are in the future. Some economists advocate particular methods of estimating these parameters based on what is allegedly a revealed preference methodology, which has been contested at length by economists, philosophers, and others (Nordhaus 2007, Fleurbaey et al. 2019).

In environmental discourse *sustainability* is frequently cited as a goal, in roughly the sense of meeting the needs of the present without compromising the ability of future generations to meet their own needs (Brundtland 1987, Solow 1991). Endorsing sustainability as the *ultimate* goal of policy would imply that policy should aim not to maximize (discounted) expected value into the future (as in standard SPA), but rather to do ‘good enough’ for future generations. SPA can be modified via an alternative SWF to encode such a sustainability objective (Fleurbaey 2015).

This section has shown that SPA has resources to account for many values at stake in challenges familiar from environmental policy and philosophy. More generally, this section has provided a conceptual overview of SPA, aimed at providing a suitable background for philosophical engagement. For more details, see leading textbooks and other resources on environmental economics, social choice and welfare, and health economics (Kolstad 2010, Adler and Fleurbaey 2016, Drummond et al. 2005).

3 CRITIQUES OF STANDARD POLICY ANALYSIS

A consequentialist view typically grounds all value in consequences for normatively relevant individuals, where the relevant individuals and the value of the consequences is explained by a theory of what is fundamentally valuable. Common views about what is fundamentally valuable include hedonism, preference satisfaction, objective list views of wellbeing (including possession of particular capabilities), and (more common in environmental philosophy) biocentric views on which being fulfilled as a living organism is fundamentally valuable. These views share the structural assumption that there is some set of individuals that fundamentally matters (e.g. on a hedonist view, those that can experience pleasure and pain), whereas other individuals do not fundamentally matter. In the environmental realm, holism is also a common view, on which ecosystems and other holistic entities have fundamental value (Hiller 2014, Sandler 2018). Consequentialist views sometimes also assign fundamental value to inequality and other properties of *distributions* of good and bad consequences to individuals as well (Adler and Fleurbaey 2016).

3.1 IS STANDARD POLICY ANALYSIS INADEQUATE IF THE PREFERENCE-SATISFACTION VIEW IS NOT ULTIMATELY CORRECT?

From the discussion in the previous section, it is clear that SPA has impressive resources to value outcomes in terms of human preference satisfaction. At the same time, given its focus on human preference satisfaction, consequentialist philosophers are often quick to argue that SPA is inappropriate for policy analysis, on the grounds that the correct consequentialist view is not the anthropocentric version of the preference satisfaction view.

However, this argument is too quick. The problem is that even if we assume for the sake of argument that the correct fundamental view is not the preference satisfaction view, it could still be true that SPA is our best option for policy analysis, and provides very reliable estimates. As an example to illustrate, human preference satisfaction as measured by SPA might be so closely correlated with an anthropocentric objective list view of wellbeing that there could be little extensional difference between what SPA implies and what an analysis using that objective list theory would imply – and furthermore, while we actually have methods for doing analysis using SPA, in contrast we arguably do not have similarly adequate methods for using the objective list theory directly. If that were all true, then from the perspective of such an objective list theory there would be no objection to the use of SPA for policy analysis, and in fact it would be a mistake not to use SPA. As one possible realistic example of this type, one might imagine an objective list theory of wellbeing in connection with policy challenges such as how best to promote the UN Sustainable Development Goals, which include goals that closely correspond to items on many objective list theories. Because there will be synergies and tradeoffs between these goals that have to be analyzed in a maximally complex system of coupled human and natural systems (Nilsson et al. 2016), and because there is arguably no better theory of how to make tradeoffs between goals on the objective list beyond what can be implemented using Enlightened SPA, policy analysis that uses Enlightened SPA might provide the best feasible estimates of policy questions such as how to best invest scarce resources to promote these goals, even assuming the truth of an objective list theory. Upon reflection, all of this should come as no surprise, given the familiar point that the best models for practical purposes sometimes rely on false simplifying assumptions about fundamental facts.

The upshot is that whether or not SPA is adequate for policy analysis in a particular context depends on whether a SPA analysis can, in that context, provide a sufficiently reliable approximation of the correct consequentialist analysis – and in many contexts, it seems that we might indeed expect Enlightened SPA to provide a sufficiently reliable approximation.

3.2 A FRUITFUL APPROACH TO IMPROVING POLICY ANALYSIS

With the preceding in mind, a more fruitful approach to improving policy analysis is not to dismiss SPA out of hand on the grounds that the correct fundamental values are not explicitly represented, but rather to focus on the question of whether SPA is good enough in particular contexts, or whether instead SPA can feasibly be improved to approximate better what ultimately matters. If we believe SPA is inferior to some feasible alternative, our critique should then aim to make such an alternative precise and readily implementable in actual policy analysis. In particular, we should characterize precisely how the relevant tradeoffs should be made by the axiology, and what decision theory should be used. If our critique lacks this level of precision, then it fails to offer any helpful recommendation for how to improve policy analysis.

As a simple example to illustrate, a theorist could argue that health should be valued more highly relative to other goods than it is in SPA, on the basis of an argument that it is more important to what ultimately matters than is indicated by human preferences alone. If this is to have relevance to policy analysis, a precise account should be offered of how health should be weighed and traded off with other goods on this view (or what range of assumptions should be used to test the sensitivity of policy recommendations). Or, as another example, a theorist could argue that ecosystem health has great fundamental value. If this is to have relevance to policy analysis, a precise formula for measuring ecosystem health should be specified in a way that is implementable in policy analysis. As these two examples indicate, different consequentialist views may generate different and sometime incompatible recommendations for how policy analysis should be modified.

In some cases, there might be broad consensus among leading normative theories as to how policy analysis should be improved. As one example from the previous section, there is presumably broad consensus among leading normative theories that Enlightened SPA is an improvement over a version of SPA that focuses merely on maximizing GDP. In the next section, we provide another example of an improvement that is possible over even best practices Enlightened SPA, related to animal welfare. This also provides a worked example of how normative theorists might aim to have positive impact on policy analysis by working directly on the details of such improvements and providing precise methods that can be readily implemented in actual policy analysis.

3.3 A FUNDAMENTAL PROBLEM WITH SPA: ANTHROPOCENTRISM

There is widespread consensus among normative theorists that an important problem with even the best anthropocentric methodology is that animals and other aspects of nature within such a methodology are always valued merely in terms of their value to *humans* (Schmidtz and Shahar 2018, Sandler 2018, McShane 2018, Palmer et al. 2014, Hiller et al. 2014, Sarkar 2012, Gruen 2011, Jamieson 2008, O'Neill et al. 2008, Ng 1995). In other words, SPA valuation is always in terms of the ultimate value of outcomes to humans only, and thus assigns no fundamental value to the wellbeing of animals.[4] For example, on even the best anthropocentric approach, the deaths of billions of birds due to climate change would have disvalue only insofar as the deaths of those birds have disvalue to humans. Most normative theorists would object that this way of valuing animal lives is fundamentally incorrect because it ignores the value of the birds' own wellbeing irrespective of its contribution to human wellbeing, as scientists and theorists broadly agree that animals, like humans, experience different levels of wellbeing depending on decisions made by others, and there is no normatively principled argument that the wellbeing of animals should be ignored while that of humans should not (Singer 1975, Kagan 2019). Further, the assumption of anthropocentrism is dubious even from within the logic of mainstream economics, since sophisticated animals have preferences over outcomes and there is nothing within economics that explains why the preferences of this subset of individuals should be ignored, just as there was never an economic logic to ignoring the preferences of people in an earlier time based on their race, gender, or other factors. As a result, normative theorists generally agree that the wellbeing of animals must be included in any full accounting of the wellbeing consequences of decisions.

[4] Similarly, anthropocentrism assigns no fundamental value to the health of ecosystems, which is a different criticism – see the references in previous sentence, and in addition Chan et al. 2016 and Dasgupta 2014 and the references therein.

3.3.1 THE CHALLENGE OF INTERSPECIES COMPARISONS

Animal welfare is almost never included in policy analysis, partly due to methodological prejudice, but increasingly also because we do not currently have good methods for quantifying animal wellbeing consequences and putting them on the same scale as quantified human wellbeing consequences. We might call this 'the challenge of interspecies comparisons'.

Recent work by Kevin Wong has highlighted the most difficult problem that needs to be solved in connection with interspecies comparisons, which is how to estimate the *wellbeing capacity* (wellbeing potential) of members of a non-human species relative to the wellbeing capacity of humans (Wong 2016). If we knew how to make those interspecies comparisons of wellbeing capacity, then we could integrate animal welfare consequences into existing methods of decision analysis, by deriving empirically-based estimates of animal welfare consequences on the same scale as human consequences that typically underpin welfarist decision making analyses.

For example, suppose an additional degree of climate change will cause us to lose 1 million life years of a particular species of bird, and we want to value this on the same scale as losses to human life from an additional degree of warming that are already modeled and valued based on an assumption about the value of one human life year. If we had a good estimate of the wellbeing capacity of that species of bird relative to a human, we could then multiply that estimate by the purely empirical impact estimate of 1 million life years lost to get an estimate of the amount of wellbeing lost by that bird species on the same value scale as the existing estimate of human wellbeing loss, assuming that one degree of additional climate change does not change the quality of life of those birds. And if one additional degree of warming does diminish the quality of life of the remaining birds of that species, we can simply the multiply the number of remaining bird species life years by a further *quality of life adjustment* term that is itself an empirical impact estimate from zoological experts and the like. (We can also use such a term to take into account any antecedent diminishment in the wellbeing experienced by all of the birds including those that would die before the warming.)

This line of thought leads to the following equation for the average wellbeing experienced by a member of a species s (which we symbolize as $\bar{u}_s$) as a function of: the average wellbeing capacity per unit of time of members of s relative to humans ($\bar{\pi}_s$), multiplied by the average duration of a life of a member of s ($\bar{\delta}_s$), multiplied by a quality of life adjustment term ($\bar{f}_s$):

(2)

$$\bar{u}_s = \bar{\pi}_s * \bar{\delta}_s * \bar{f}_s$$

The key point here is to highlight the term $\bar{\pi}_s$ as the key unknown term, where the unsolved problem of how to estimate $\bar{\pi}_s$ is the essence of the challenge of interspecies comparisons. (The other terms $\bar{\delta}_s$and $\bar{f}_s$ are susceptible to existing empirical methods, where the term f can be seen as the focus of existing animal welfare science – see for example Browning 2019, Appleby et al. 2011, and Fraser 2008.[5])

3.3.2 A METHOD FOR QUANTIFYING ANIMAL WELFARE AND MAKING INTERSPECIES COMPARISONS

[5] Compare also the term $\bar{f}_s$ to McMahan 2002's concept of *fortune*, a connection Wong 2016 notes.

In this section we propose a method of making interspecies comparisons that has some analogy to the method used in Equation 1 of taking consumption as a proxy for human wellbeing: the proposal is to make interspecies comparisons based on a proxy that is imperfect but yet delivers estimates as good as we can expect in practice. To do this we first identify a proxy, call it n, to use as the basis for estimating wellbeing potentials across species, analogous to the use of consumption (c) above as the basis for estimating wellbeing across humans. As an overly simplistic illustration of this idea, n might be the number of neurons in the brain of members of a species. Data on number of neurons is readily available, and may be a good proxy in some select contexts, such as 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'. When greater accuracy is required for specific species or individuals, n can be set equal to a more complex metric based on expert analysis of empirical properties that are best correlated with different levels of wellbeing (which might differ according to different substantive theories of wellbeing) – e.g. the number of neocortex-like neurons, cortisol levels, sociality, or other leading factors identified by the scientific community and philosophers as most closely correlated with wellbeing capacity (Herculano-Houzel 2017, Tye 2017, Olkowicz et al. 2016, Klein and Barron 2016, Barron and Klein 2016, Shriver 2014, Dawkins 2012, Appleby et al. 2011, Fraser 2008).

Abstracting from those details, which are not essential to the core challenge of how to make interspecies comparisons, the first step of the proposal is to parameterize such an empirical proxy n, perhaps with an exponential weight ψ, into estimates of comparative wellbeing capacity for different species. The second step is to multiply this estimate of wellbeing capacity by a descriptive measure of the degree to which this potential is actually realized, and multiply by the f quality of life and δ duration terms, to yield the desired wellbeing estimates. For example:

(3)

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

(In ordinary language: the total sum of wellbeing is approximately equal to the sum over all individuals across species of: that individual's empirical basis for wellbeing capacity raised to the normative exponent (which determines the relationship between the empirical proxy and wellbeing capacity) multiplied by the quality of life adjustment, multiplied by the duration of that individual's life.)

In practice, it would often be more feasible without important loss of accuracy to use species-level averages (where averages are denoted by a bar over the letter) as the proxy for wellbeing potential, which can then be multiplied by the species population P_s:

(4)

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

Equations 3 and 4 summarize the 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 the normative parameter such as ψ that are grounded in normative and empirical considerations (on analogy with how values for θ in Equation 1 is grounded in normative and empirical considerations), and empirically-determined values for f and δ.

Note that equation 4 provides a practical method of estimating the value of the following more theoretically obvious equation that multiplies the population P_s of each species by the average wellbeing $\bar{u}_s$ of members of that species:

(5)

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

The problem of interspecies comparisons means that we cannot directly use equation 5 prior to a method of making interspecies comparisons such as that developed above, as using equation 5 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 method for making those comparisons, such as provided by equations 3 and 4: namely, to take $\bar{u}_s \;\approx\; \bar{n}_s^{\psi} \bar{f}_s \bar{\delta}_s$.

Note that if we were to translate Equation 2 above into a total utilitarian axiology, this would yield:

(6)

$$W^{TU} \approx \sum_s P_s \, \bar{\pi}_s \bar{f}_s \bar{\delta}_s$$

When we substitute the term $\bar{n}_s^{\psi}$ for $\bar{\pi}_s$ in Equation 6, the result is Equation 4 above. The term $\bar{n}_s^{\psi}$ can similarly be incorporated into other population axiologies in a straightforward way, but for ease of exposition we focus only on totalism in this chapter. (See Budolfson and Spears 2018, forthcoming, and 2019b for further discussion of population ethics, including in connection with animal welfare.)

3.3.3 ESTIMATES OF OPTIMAL TRADEOFF RATES BETWEEN HUMANS AND ANIMALS, AND THE REPUGNANT CONCLUSION

Figure 1 illustrates how a sensitivity test could be incorporated into policy analyses based on different principled ways of using the parameter ψ to estimate potential wellbeing 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/Amph	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 wellbeing (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 can be used to put human life years (which can be estimated via familiar proxies such as equations 2 or 3 above) on the same scale as the life years of animals of different species, and each estimate does so in a principled way that is empirically grounded. For example, assuming number of neurons as a basis for wellbeing estimates, if ψ is set equal to 2 (a principled lower value for animals), then a human life year is worth almost 120,000 mammal life years, and almost 120,000,000 fish life years. If instead ψ is set equal to 1 (a principled higher value for animals), then a human life year is worth about 344 mammal life years, and about 10,700 fish life years. These alternative estimates appear to represent much of the range of empirically-grounded and principled views over the wellbeing of animals of different species (Herculano-Houzel 2017), and can avoid unintuitive implications. It may not even be desirable to attempt to choose between these estimates in policy analysis, if the goal is to take normative uncertainty into account and test the sensitivity of optimal decisions to this range of different reasonable (and empirically and theoretically principled) estimates.

Of particular note are the implications (and lack thereof) for the repugnant conclusion. One might think a priori that assigning any positive value to the lives of insects could dominate analyses in a repugnant conclusion-like way because there are many quadrillions of insects (Singer 2016; see also Tännsjö 2016, Fischer 2016). For a real-world example, insects are estimated to benefit in numbers and in average wellbeing from climate change, and so there is some worry that if they are assigned any positive value, that will imply that we should do nothing about climate change (Sebo forthcoming).

However, using the valuations above, it can be shown that a repugnant conclusion need not be an implication of assigning positive value to insects in a principled way. For example, the method in the previous section provides a way of demonstrating that this need not be true for principled valuation scheme with ψ = 2, which assigns value to insects using the same principled 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. When these higher ψ values are used in climate policy analysis, insects do not dominate the calculation because they are assigned such miniscule value in a principled way by the ψ exponent (Budolfson and Spears 2019a). Still, under $\psi = 1$ (a higher valuation of animals that is uncommon in general society), it is true that insects can dominate the calculation because of their sheer numbers together with that higher valuation of their lives. So, adding valuation of animal wellbeing in a principled way does not necessarily lead to a repugnant conclusion if insects are given positive wellbeing in a principled way (e.g. ψ = 2), but insects can indeed 'repugnantly' dominate at higher valuations of 'lower' animals (e.g. ψ = 1). This is an instance of the general possibility for every approach to tradeoff-making social evaluation to yield repugnant-seeming outcomes when applied to large numbers (Budolfson and Spears 2018).

In sum, the method outlined here allows interspecies comparisons based on empirically-available estimates of species population dynamics and within-species quality of life adjustment, together with empirical proxies for wellbeing capacity n that can be calibrated with the ψparameter to reflect normative uncertainty about the connection between those empirical proxies and wellbeing capacity. Implementing these methods in policy analyses would have an important impact on estimates of how best to invest time and money by individuals (Budolfson 2015), businesses (Berkey forthcoming), and charities (including for purposes of 'effective altruism') (Wong 2016, ACE 2019, OPP 2019), and similarly for estimates of optimal public policies for correcting market failures (Cowen 2006, Norwood and Lusk 2011, Jarvis and Donoso 2018), for sustainable intensification of agriculture that aims to take animal welfare into account (producing more food while reducing the overall impacts of agriculture) (Garnett et al. 2013), for climate change policy (how quickly we should be reducing greenhouse gas emissions) (Hsiung and Sunstein 2007, Budolfson and Spears 2019a), and for wilderness protection policy and other challenges related to natural resource management (Hsiung and Sunstein 2007, Sunstein 2018 chapter 6, Fischer et al. forthcoming). In all of these cases, if the wellbeing of animals is taken more fully into account, then decisions by individuals, governments, and others will become better on welfarist grounds.

4. CONCLUSION: A PERSPECTIVE ON CONSEQUENTIALISM AND POLICY ANALYSIS

Enlightened SPA is a powerful first step toward correct consequentialist policy analysis. However, it is not capable of including fundamental valuation of the wellbeing of non-human animals, and thus must be supplemented with methods for including their wellbeing, at least if the goal is to provide a philosophically defensible form of welfarist policy analysis. Many other important objections were briefly noted above or in some cases not discussed in detail: for example, some would argue that welfarist policy analysis should be constrained by fairly strict ‘side constraints’ to respect basic human rights and less stringent constraints to respect the outputs of institutions necessary for wealth creation including property rights, corrective justice, and free exchange. These constraints could themselves be justified on consequentialist grounds as optimal features of the basic structure of society. Perhaps policy analysis should also be constrained by extreme modesty about the limitations of impact assessments to include unintended consequences and the likelihood of endogenous solutions not comprehended by policy projections.

A more controversial perspective is that together, these constraints create a default toward a form of classical liberalism, where this default is often overridden in particularly clear cases such as air pollution regulation and climate change, or in the provision of basic public goods, including funding basic research on medicine and public health, and perhaps a social minimum of resources needed for a healthy life (Budolfson 2017). When the default is overridden and regulation is needed, (constrained) welfarist consequentialism is appropriate for policy analysis, implemented via advanced social welfare functions of the sort advocated here, which can capture at least most dimensions of value in most cases, including for animal welfare, valuing inequalities and other considerations of distributive justice, and thus provide the best tools for realistic policy analysis in complex societies.

A key remaining question is what the correct parameters are to use in these social welfare functions to value aversion to inequality, the comparative wellbeing capacity of humans and other species, and other key parameters for determining social aggregation along which there is currently normative uncertainty. These are key issues for further research. By working directly on the details of calibrating these parameters, and delivering precise proposals for other improvements that can be readily implemented in policy analysis, normative theorists can aim to have positive impact on policy analysis.

REFERENCES

Adler, Matthew, and Marc Fleurbaey eds. 2016. *Oxford Handbook of Well-Being and Public Policy*. Oxford UP.

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

ACE 2019. Animal Charity Evaluators, https://www.animalcharityevaluators.org/.

Arrow, Kenneth and Anthony Fisher. 1974. Environmental Preservation, Uncertainty, and Irreversibility, *The Quarterly Journal of Economics* 88(2): 312–319.

Banzhaf, H.S. 2009. The Political Economy of Environmental Justice, *Resources*: 25 May 2009.

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.

Berkey, Brian. forthcoming. Prospects for an Animal-Friendly Business Ethics, in Thomas ed. *Animals and Business Ethics*, Springer.

Broome, John. 2012. *Climate Matters*. Norton.

Browning, Heather. 2019. *If I Could Talk to the Animals: Measuring Subjective Animal Welfare*, PhD Dissertation, Australian National University.

Brundtland, Gro. 1987. *Report of the World Commission on Environment and Development: Our Common Future*. United Nations General Assembly document A/42/427.

Budolfson, Mark. 2012. Collective Action, Climate Change, and the Ethical Significance of Futility. Working paper, http://www.budolfson.com/papers.

---------- 2015. Consumer Ethics, Harm Footprints, and the Empirical Dimension of Food Choices, in Chignell, A., Cuneo, T., and Halteman, M. eds. *Philosophy Comes to Dinner*, Routledge.

---------- 2017. Market Failure, the Tragedy of the Commons, and Default Libertarianism in Contemporary Economics and Policy, in David Schmidtz and Carmen Pavel eds. *Oxford Handbook of Freedom*, Oxford UP.

---------- 2018. The inefficacy objection to consequentialism and the problem with the expected consequences response. *Philosophical Studies* 176: 1711-1724.

Budolfson, Mark and Dean Spears. 2018. Why the repugnant conclusion is inescapable. Working Paper, Princeton CFI.

---------- 2019a. Optimal climate policy including animal welfare. Working paper, Princeton CFI.

---------- 2019b. Methods for quantifying animal wellbeing and estimating optimal tradeoffs against human wellbeing — and lessons for axiology, including new arguments for separability. Working paper, Princeton CFI.

----------- 2019c. An impossibility result for decision-making under normative uncertainty. Working paper, Princeton CFI.

---------- forthcoming. Does the Repugnant Conclusion have important implications for axiology or for public policy?, in Gustaf Arrhenius, Krister Bykvist, and Tim Campbell eds., *Oxford Handbook of Population Ethics*, Oxford UP.

Buchak, Lara. 2013. *Risk and Rationality*. Oxford UP.

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

Connelly, Matthew. 2008. *Fatal Misconception*. Harvard UP.

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.

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

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

Drummond, Michael et al. 2005. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd edition. Oxford UP.

Ellerman, Denny et al. 2000. *Markets for Clean Air*. Cambridge UP.

Fischer, Robert. 2016. What if Klein & Barron are right about insect sentience? *Animal Sentience* 9(8).

Fischer, Robert, Christian Gamborg, Jordan Owen Hampton, Clare Palmer, and Peter Sandøe, forthcoming. *Wildlife Ethics*, Wiley.

Fleurbaey, Marc. 2015. On Sustainability and Social Welfare, *Journal of Environmental Economics and Management* 71: 34–53.

Fleurbaey, Marc et al. 2019. The social cost of carbon: valuing inequality, risk, and population for climate policy, *Monist* 102: 84–109.

Fraser, David. 2008. *Understanding Animal Welfare*. Wiley.

Garnett, Tara et al. 2013. Sustainable Intensification in Agriculture: Premises and Policies, *Science* 341 (6141), 33-34.

GiveWell.org. 2019. Overview. https://www.givewell.org/about/givewell-overview.

Gruen, Lori. 2011. *Ethics and Animals*. Cambridge UP.

Hausman, Daniel, Michael McPherson, and Debra Satz. 2018. *Economic Analysis, Moral Philosophy and Public Policy*. 3rd edition. Cambridge UP.

Herculano-Houzel, Susan. 2017. Numbers of neurons as biological correlates of cognitive capability, *Current Opinion in Behavioral Sciences* 16: 1-7.

Hiller, Avram. 2014. System Consequentialism, in Hiller et al. eds. *Consequentialism and Environmental Ethics*. Routledge.

Hiller, Avram, Ramona Ilea, and Leonard Kahn eds. 2014. *Consequentialism and Environmental Ethics*. Routledge.

Hsiung, Wayne and Cass Sunstein. 2007. Climate Change and Animals, 155 *Univ. of Pennsylvania Law Review*. 1695.

IPBES. 2019. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. https://www.ipbes.net/.

IPCC. 2014. United Nations Intergovernmental Panel on Climate Change. https://www.ipcc.ch/.

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

Jarvis, Lovell and Pablo Donoso. 2018. A Selective Review of the Economic Analysis of Animal Health Management, *Journal of Agricultural Economics* 69(1): 201–225.

Kagan, Shelly. 2011. Do I make a difference? *Philosophy & Public Affairs*: 39, 105–141.

---------- 2019. *How to Count Animals*. Oxford UP.

Kahneman, Daniel and Angus Deaton. 2010. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences* 107(38): 16489-16493.

Klein, Colin and Andrew Barron. 2016. Insects have the capacity for subjective experience. *Animal Sentience* 9(1).

Kolstad, Charles. 2010. *Environmental Economics*. 2nd edition. Oxford UP.

Lam, David. 2011. How the World Survived the Population Bomb: Lessons From 50 Years of Extraordinary Demographic History. *Demography* 48(4), 1231-1262.

MacAskill, Will, Krister Bykvist, and Toby Ord. forthcoming. *Moral Uncertainty*. Oxford UP.

McMahan, Jeff. 2002. *The Ethics of Killing*. Oxford UP.

McShane, Katie. 2018. Why Animal Welfare Is Not Biodiversity, Ecosystem Services, or Human Welfare, *Les Ateliers de l'Éthique/The Ethics Forum* 13(1): 43–64.

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

Mikati, Ihab et al. 2018. Disparities in Distribution of Particulate Matter Emission Sources by Race and Poverty Status, *American Journal of Public Health* 108(4): 480-485.

Nefsky, Julia. 2011. Consequentialism and the problem of collective harm: A reply to Kagan. *Philosophy & Public Affairs*: 39, 364–395.

---------- 2019. Collective Harm and Individual Inefficacy, *Philosophy Compass* 14 (4): e12587.

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

Nilsson, Mans, Dave Griggs, and Martin Visbeck. 2016. Policy: Map the interactions between Sustainable Development Goals, *Nature* 534: 320-322.

Nolt, John. 2011. How Harmful Are the Average American's Greenhouse Gas Emissions?" *Ethics, Policy and Environment* 14(1): 3-10.

Norcross, Alastair. 2004. Puppies, pigs, and people. *Philosophical Perspectives*: 18, 229–245.

Nordhaus, William. 2007. A review of the Stern review on the economics of climate change." *Journal of Economic Literature* 45(3): 686-702.

Norwood, F. Bailey, and Jayson Lusk. 2011. *Compassion by the Pound*. 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.

OPP. 2019. Open Philanthropy Project, Farm Animal Focus, https://www.openphilanthropy.org/focus/us-policy/farm-animal-welfare.

Otte, M. Joachim and Pius Chilonda. 2000. Animal Health Economics, in Livestock Information (Rome, Italy: Sector Analysis and Policy Branch, Animal Production and Health Division (AGA), FAO.

Ostrom, Elinor. 1990. *Governing the Commons*. Cambridge UP.

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

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

Portmore, Douglas. 2011. *Commonsense Consequentialism*. Oxford UP.

Sandler, Ronald. 2018. *Environmental Ethics*. Oxford UP.

Sarkar, Sahotra. 2012. *Environmental Philosophy*. Wiley-Blackwell.

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

Scovronick, Noah 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. *Philosophy and Climate Change*, Oxford UP.

Shrader-Frechette, Kristin. 2002. *Environmental Justice*. Oxford UP.

Shriver, Adam. 2014. The Asymmetrical Contributions of Pleasure and Pain to Subjective Well-Being, *Review of Philosophy and Psychology* 5:135-153

Singer, Peter. 1972. Famine, Affluence, and Morality, *Philosophy & Public Affairs* 1(3): 229-243.

---------- 1975. *Animal Liberation*. HarperCollins

---------- 1980. Utilitarianism and vegetarianism. *Philosophy & Public Affairs*: 9, 325–337.

----------- 2016. Are insects conscious?, *Project Syndicate*.

Solow, Robert. 1991. *Sustainability: An Economist's Perspective*. The 18th J. Seward Johnson Lecture. Woods Hole.

Stiglitz, Joseph, Amartya Sen, and Jean-Paul Fitoussi. 2009. Report by the commission on the measurement of economic performance and social progress. France.

Sunstein, Cass. 2014. *Valuing Life*. Univ. of Chicago Press.

---------- 2018. *The Cost-Benefit Revolution*. MIT Press.

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

Tye, Michael. 2017. *Tense Bees and Shell-Shocked Crabs*. Oxford UP.

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