

Explicit Methodologies for Normative Evaluation in Public Policy, as Applied to Carbon Budgets

KIAN MINTZ-WOO 

ABSTRACT *What could philosophical or justice perspectives contribute to climate (and other applied philosophy) policy discussions? This question is important for philosophers on government policy committees. This article identifies two novel concerns about such contexts (which I call ‘contingent selection’ and ‘committee deference’) and systematizes some potential methodologies before arguing for a previously unrecognized methodology that focuses on disciplinary convergence. After supporting this methodology by providing several justifications, the Appendix explains how to apply it when evaluating a carbon budget. This methodology and its application are meant to be of interest to both theorists and practitioners.*

1. Introduction

Carbon budgeting is an increasingly common tool in the development of just transitions and climate policy.¹ But when would a carbon budget be just (or justified)?² What role is appropriate for philosophers to play on policy committees addressing such matters? More practically, what methodology or methodologies could be appropriate for adding normative evaluation (including, but not limited to, applied ethics, moral philosophy, and political philosophy) to such complex policy questions?

I think these questions are intrinsically important and interesting, but answering them can also be motivated for extrinsic reasons: since non-philosophers are uncertain about what philosophers do and how we decide, having an explicit methodology in the policy process is especially important.

To level set context, suppose that a philosopher finds herself on a committee tasked with normatively evaluating and recommending some particular policies in some particular applied area (e.g. on a climate committee, proposing national carbon budgets). In some such contexts, her role will be limited by legal and political constraints, but to the extent that this is not the case, what role should she adopt? Is there a role that would take advantage of her philosophical competences?

I will introduce two novel challenges that are important for her to consider. The first I call ‘contingent selection’. The point is that her selection to the committee (for instance, by some mid-level governmental actor) is subject to all kinds of contingencies. The result is that, from the point of view of the policy process, her particular philosophical or normative positions could be objectionably *arbitrarily selected from wide variation* or, closely relatedly, *non-representative* of the discipline as a whole.³

The second I call ‘committee deference’. Assuming that she finds herself among relative philosophical naïfs, whatever substantive position she puts forward or defends as philosophically respectable, that view is likely to receive deference as philosophically

respectable. The result is that, regardless of her substantive views, she will *command agreement*.

I argue that the two dominant views about her philosophical role implicit in the literature are open to objections, including on the grounds of these challenges. These positions hold that her role is to develop and apply the normative position she believes is strongest ('the partisan') or to systematize and apply the normative position she believes to reflect society ('the populist').⁴

The alternative view I defend is that her role is to act as a representative of the discipline, by giving some broad range of relevant philosophical positions, together with an indication of where they converge (or diverge) on the particular policy question at hand ('the convergent evaluator'). When playing this role, her own philosophical judgments are less material or even immaterial.

In Section 2, I introduce these two novel challenges: contingent selection and committee deference. In Section 3, I critically introduce two standard roles that she might play ('the partisan' and 'the populist'). In Section 4, I explain and defend the third account ('the convergent evaluator'). In Section 5, I consider an objection from the concern that the philosophers' views might diverge from the public. In Section 6, I conclude. The Appendix A includes a carbon budget case study where these arguments are applied based on the author's experience with such policy contexts.

2. Two Novel Challenges

Consider a philosopher tasked with contributing to a policy committee. The committee may have political representatives, representatives of civil society, including some mix of thinktank and academic experts. They are tasked with (individually) making evaluations and (jointly) recommending some policy proposal. Let us suppose, as I imagine is most common, that the philosopher is the lone philosopher (perhaps presented as 'the ethicist' or 'the justice theorist') – among few others with even a passing familiarity with philosophy.

I will assume that the selection of a philosopher presumes that there is *something* the committee expects the philosopher *qua* philosopher to add, even if they are unsure exactly what this is. (If she is just appointed as merely another intelligent committee member, then it is plausible there is no special role she should adopt. However, in this case, she would presumably *not* be 'the ethicist' or 'the justice theorist').

The committee members are charged with considering and evaluating some candidate policies and, subsequently, to jointly debate with the ultimate aim of recommending some policy to a policy-maker. Recommendation in hand, the policy-maker may adopt, adjust, or reject the committee's view.

This committee context is susceptible to two challenges that I believe have not been recognized and which are fraught morally and from the point of view of political theory. They suggest that the role a philosopher should play on such a committee might be quite different from the role a philosopher plays in most of her working life.⁵

In Section 2.1, I introduce the challenge I call 'contingent selection', and in Section 2.2, I introduce the challenge I call 'committee deference'.

2.1. Challenge 1: Contingent Selection

Committees are appointed with various invited experts. Generally speaking, the appointers are governmental employees, and they cannot be expected to have expertise in all (or even any!) of the areas that the experts they invite do. Even when they are familiar with the relevant domains, they are almost by definition less informed than those they invite. This is not by itself always a major problem, but it can become one if we can expect there to be a wide range of positions the expert might adopt which the appointers do not or cannot distinguish between in the expert's selection. In short, if selecting an expert, will the expert's evaluation be *representative* of the positions and state of knowledge among the expert's community – or at least highly likely to be representative? Alternatively, is there a wide range of variation among the evaluations that could result – such that it is unclear or practically inaccessible for the appointer to appreciate *ex ante*?

I believe that the challenge of *contingent selection* is far more serious or difficult to address in the context of normative policy advice than for most scientific types of policy advice. To illustrate, consider a carbon budgets committee that is looking for a climate scientist. On fundamental questions of climate change, such as the commitment to a significant anthropogenic role, there is widespread agreement.⁶ In contrast, on fundamental questions of normative approaches to climate change, there is no similar 'mainstream' normative theory with philosophers having comparable distributions among, say, virtue ethicists, deontologists, and consequentialists.⁷ It follows that, in selecting a climate scientist, one is overwhelmingly more likely to get similar scientific explanations or assessments, since there is a broadly shared fundamental base, regardless of whether one is antecedently familiar with that particular scientist's views. However, when selecting a philosopher, if one is not *very* familiar with that philosopher's particular views, there is significant variation on both the fundamental questions (e.g. what aspect of the moral world is of crucial importance: virtues or obligations) and on their policy implications (e.g. consequentialists might be more inclined to engage in carbon pricing policies whereas virtue ethicists might be interested in messaging campaigns and developing green norms) which could influence her evaluation.

One might object that there is an important disanalogy: climate scientists are not experts in *policy* evaluation, so if we were to ask them to engage in policy evaluation, they may well have a similarly broad variation of views. While this is true, it is not a relevant disanalogy because the question is what competence they bring to the policy committee. The climate scientist's role should be to explain, based on current scientific understanding, what various policy proposals are likely to mean in climatic terms. The normative philosopher's role should be to evaluate policy proposals with respect to normative or moral considerations. My claim is that, restricted to their relevant domains, most climate scientists are not likely to vary greatly from their colleagues on these fundamental questions, whereas normative theorists do vary greatly in their normative positions.

Of course, not all such selections are as uninformed as this challenge presents; sometimes, committee appointments are made because the appointers want a particular position represented – or, indeed, a particular style of thought (e.g. theological as opposed to secular; consequentialist as opposed to rights-based). While I believe this happens less than most philosophers would expect, when it occurs it might have the surprising effect of *supporting* my subsequent arguments. If a philosopher suspects she is being selected for her substantive normative position, she has reason to avoid simply applying her position (*à la*

‘the partisan’) and to provide a less idiosyncratic analysis (*à la* ‘the convergent evaluator’). The point is that she may have reason for wanting to avoid the kind of normative bias that an appointer might wish to introduce.⁸

2.2. Challenge 2: Committee Deference

In many committee cases, a philosopher is lucky to find herself included – most policy committees lack philosophers – so, if she is included, it is not unlikely that she will be alone. If she is even more lucky, other members of the committee may have taken courses in ethics or philosophy at some point. Even if members have a keen interest in philosophy, the specific applied philosophical arguments and positions relevant to the policy matter at hand are unlikely to be familiar to them – certainly nowhere as clear as to the professional philosopher working in this area.

The curse of knowledge is relevant here: it is difficult for a seasoned philosopher to recognize all that she knows, both theoretically and also about the variety of positions being discussed and the subtle arguments advanced for and against these. In general, it will be hard for her to remember or understand how much background information she has that her fellow committee members lack. When I participated in such a committee, even some of what philosophers would take to be basic terms of the relevant debates were completely novel ways of framing the problems for my fellow committee members (for instance, that emissions are not intrinsically morally valuable – they are instrumentally valuable insofar as they can help secure improved wellbeing).

There are two flavours of *committee deference* that I would distinguish: *committee deference based on representation* and *committee deference based on plausibility*. First, unless the philosopher repeatedly caveats her claims as her positions or as contested positions, the committee will neither know when the views expressed are her own nor know whether they are contested. The result is that we can expect that the committee will take her claims to be generally accepted by philosophers. On the one hand, this is a problem because they may take various controversial positions as uncontroversial. On the other hand, this is a problem because they may not recognize the distinction between claims that *do* command broad acceptance and those that are controversial.

For instance, if a climate ethicist explains and endorses the beneficiary pays principle without explaining alternative principles (and without labelling her principle as contested), the committee might reasonably (but incorrectly) assume that climate ethicists broadly accept this principle. Philosophers are not trained to caveat claims in this way because they are usually talking with specialists who are familiar with the debates and alternative positions, so this is an issue that is likely to occur unless the philosopher is especially careful.

Second, the philosopher is trained in making plausible arguments and giving strong defences. For philosophical naïfs, regardless of the substantive position, a philosopher defending the view she endorses is likely to be highly convincing. Thus, we can expect the committee to find the philosopher’s views plausible regardless of how plausible they would appear to another professional philosopher. There are a few reasons for this, but the main ones are that the committee will not have been exposed to the various positions, so they will not easily understand the alternatives, and also that weaknesses in those positions that philosophers are trained to spot will not be as obvious to the committee.

In short, both because the philosopher may be taken by the committee as having representative and uncontested views (in the absence of appropriate caveats) and because her arguments are likely to appear highly plausible to the committee, the committee is likely to defer to and believe the philosopher's substantive conclusions are philosophically plausible. The result, I believe, is that the committee can unreasonably come to the conclusion that the particular normative judgments of the philosopher are 'what philosophers believe', 'what philosophy demands', or even 'what justice or morality demands', even when those views are relatively weak or idiosyncratic.

3. Extant Methodologies

In the context of public policy, there is a small but important selection of cases where normative philosophers have participated in evaluation and recommendation processes⁹ and, perhaps even more illuminatingly, also cases where philosophers have reflected upon the appropriate role for philosophers engaged in these processes.¹⁰ While my interest is in climate policy, almost all of these are in biomedical policy questions. However, the committee paradigm is broadly comparable regardless of policy area.

With the notable exception of Wolff, few of them explicate a particular procedure that captures their methodology.¹¹ However, many of them are admirably explicit about what they take the role of a philosopher in such contexts to be (or not to be). In my reading, when it comes to evaluating and recommending policies (which I take to be distinctly normative tasks which require specific value commitments), there are a couple of families of approaches that surface.

With the key caveat that the following are not intended to be exclusive (and are certainly not exhaustive), I think it worthwhile to distinguish these approaches depending on the philosophical role that the philosopher takes on the committee: a proponent of her own considered judgment ('the partisan') (Section 3.1), or a systematizer and proponent of the values of the society she belongs to or operates within ('the populist') (Section 3.2). In both cases, I also provide a candidate procedure reflecting the role. As I discuss them, I indicate how challenges, including those noted above, can undermine these roles.

3.1. *The Partisan's Evaluation*

The most obvious methodology draws on the philosopher's own considered philosophical position. After all, she was invited to the committee – and she would appear to have the most moral expertise in the group. Her work, it is plausible to claim, is to consider the policy problem carefully, bring her knowledge of moral theories and principles to bear on it, and develop a careful, but forthright, position.

Indeed, some philosophers have even taken this role to be *unavoidable* when it comes to public policy debates on committees. For instance, Brock, who served on a presidential ethics committee, suggests that philosophers may end up pressing 'their own views on [uncertain] moral questions. [Indeed] they will *inevitably* do so, even if neither they nor others construe their role to be provider of the moral truth and of solutions to moral problems'.¹²

A simple version of this view has the philosopher make explicit value judgments and then advocate for those judgments in the relevant policy process. Momeyer defends this

view, arguing that doing so refutes (by example) the bugbear that philosophy need be wishy-washy or indeterminate.¹³

However, this simple version straightforwardly suffers from the challenges of contingent selection and committee deference. If the philosopher simply advocates her own position, from the point of view of the policy process, the resulting evaluation will be sensitive to which position she holds, and it is unlikely that she was selected because that was her specific position – rendering the evaluation subject to significant arbitrary or contingent variation from the point of view of the policy process. Furthermore, the committee may not understand in which respects her value judgments differ from her interlocutors. Finally, as she is a trained philosopher, the committee will have trouble pointing out any gaps or flaws in her position, making her substantive positions seem more plausible than may be warranted. Thus, unless she is very careful, the committee is likely to defer to her and treat her position as the philosophical truth or consensus.

One response is to make a more sophisticated version of this view whereby the philosopher explicitly identifies some alternative positions – and explains why she believes her own position to be superior. Kamm seems to adopt such a sophisticated version. She says that when there is a philosophical dissensus, and the philosopher knows that his view is contested, he ‘should present reasons as best he can for what he thinks is correct. So, while he may be bound to inform about others’ views, he is not their agent’.¹⁴ In short, the philosopher should let others know that the view is contested, explain what other philosophers have said, and then give a considered philosophical judgment: ‘the philosopher should truly philosophize, rather than merely report the views of others’.¹⁵

The good news is that this avoids committee deference based on representation, since the philosopher is careful and explicit about disagreement, so the committee understands which positions are idiosyncratic. The bad news is it does not help with contingent selection or with committee deference based on plausibility. First, her considered philosophical judgment in an explicitly contested context will vary (and may vary greatly) from those of her interlocutors. Second, her considered philosophical judgment will likely still be convincing to the committee, since she will give her arguments for her position and there is no sophisticated philosophical interlocutor present to challenge them.

Wolff discusses a procedure which we can adjust to conform to this partisan role so I include his here, with slight modifications:¹⁶

1. Identify the issue and current state of affairs
2. Identify the arguments and values*
3. Consider history and comparisons
4. Create a profile of possible solutions
5. Evaluate options with the philosopher’s considered normative judgments*

(The asterisks indicate steps that vary between the different procedures. In private correspondence, he has told me that these ‘steps’ are not necessarily sequential. I am happy to leave open whether or to what extent these steps should be considered sequential).

Since my criticisms of the partisan do not hang precisely on the details of this procedure, I refer the interested reader to his text for more explanation. The key for Wolff’s ‘engaged political philosophy’ is that investigation into the policy’s circumstances should come before any theorizing. In particular, before the philosopher begins to evaluate the policies

or employ her theoretical or philosophical frameworks, she deeply engages with the facts and practicalities in order to be appropriately sensitive to historical successes or failures.

But the key for our purposes here is that, like all versions of the partisan's evaluation, this partisan procedure takes the *philosopher's own* normative position and judgments as primary.

3.2. *The Populist's Evaluation*

Many theorists are concerned about the partisan's evaluation; some of these are concerned because they are sceptical that the philosopher *qua* philosopher has better values or more reliable access to substantive moral truths. For instance, some¹⁷ say that 'expertise about an ethical view' does not imply 'expertise in ethics' while others¹⁸ ask – when there is disagreement about a philosopher's normative claims (when is there not?) – '*How* does he know what is right?'

One response to these concerns is to endorse a democratic instinct: if the philosopher's philosophical judgments are not inherently more reliable than anyone else's, then we should consider everyone's, or at least some subset of commonly accepted normative positions. A methodology of this type would draw on the values of society.¹⁹ The associated role, then, of the philosopher is to aggregate, clarify, and systematize these values using her philosophical skillset. With a normative position built out of or reflective of this systematization of societal values, the philosopher can then apply this to the policy question in hand. This populist role, naturally, avoids the concerns associated with taking the normative position that the philosopher herself finds most plausible.

Key exemplars of this type of analysis could include Mary Warnock, who was certainly (albeit not exclusively!) sensitive to the normative positions of the public at large. Mary Warnock was an early pioneer philosopher in taking a role in a commission to recommend public policy. Of course, she did not take 'public outrage ... as a reliable guide to what is wrong, or, critically, what makes for good law'.²⁰ But she thought that 'public philosophy' needs to be at least sensitive to what is politically or democratically acceptable. So, while her view may well have been hybrid, at least one prominent source of values in the policy process for her was society.²¹

If the philosopher were to adopt this populist role, the result would mostly avoid the challenges noted in Section 2. First, in terms of contingent selection, we can expect that a philosopher who is responding to societal values will not make an evaluation that varies greatly from other philosophers. While the ways that one philosopher might systematize societal values could differ from another, the difference between the two is more constrained compared to the difference between fundamentally different moral frameworks. Second, in terms of committee deference, the philosopher is not relying on (or is relying much less on) substantive arguments for the particular normative judgments being rendered. Instead, they are representing the judgments that they think reflect the public.

Wolff's view is closest to that of a populist (he thinks that our policy judgments should surface and then reflect the deep values in society), so we can construct a populist's procedure that is very similar to Wolff's 'engaged political philosophy',²² while being explicitly sensitive to society:

1. Identify the issue and current state of affairs
2. Identify the arguments and values*

3. Determine the relevant values of society or the public, perhaps trying to render these values more systematic or coherent*
4. Create a profile of possible solutions
5. Evaluate options with reference to societal values*

Again, regardless of the specific procedure, the key to a populist's analysis is a primary appeal to *societal* value judgments.

I believe a dilemma can be offered to the populist: either the relevant value judgments are some implicit or sophisticated deep (*à la* Wolff and Walzer) values, for example, those implied by founding documents or historical commitments (in which case the specific implicit or historical source of those values needs to be determined and defended) or the relevant value judgments are those expressed or revealed by the current populace (in which case, as below, I think two objections could be offered).²³

I think there are two different objections to this populist role for the philosopher. The first is very familiar from the literature: it threatens an *objectionable conventionalism* with respect to normative conclusions. As O'Neill puts it, results of 'public opinion polls ... cannot settle ethical or policy questions, and can make at best a limited contribution to public policy formation'.²⁴ Kamm points out that this method is 'suspect since it might lead to the refusal to recommend rights for blacks or women in racist or sexist societies'.²⁵ In short, the views of society are not thought to be dispositive with respect to normative conclusions.

The second objection is that this role is *unfit for a philosopher*. Given that a populist's evaluation relies on gathering and understanding the values of society, it is not clear that a philosopher is especially well placed to engage in them. It may well be that philosophers are good at systematizing views, but the other key step in this role is determining which values are endorsed in society. It seems that understanding societal values is better done by social scientists like sociologists and psychologists. If moral philosophers are arguably less well equipped to take on these empirical matters, adopting a populist's role may not behave our committee philosopher either.²⁶

4. The Convergent Evaluation

What do philosophers add to a policy discussion? A key contribution, I believe, is understanding what inferences validly follow from normative starting points. Not only do philosophers endorse (or consider) various normative positions, but they work out in subtle detail the *implications* of and *what follows from* those positions more carefully and completely than non-philosophers (a natural formulation of this is that they *produce arguments*). Thus, a philosopher who is up-to-date on the developments or applications of a normative position is likely to have a far broader and deeper *understanding* of the strengths and weaknesses of that position as well as how it is relevant for practical applications.²⁷ They will, for instance, be able to construct or recall counterexamples to versions of that position and have various strategies for avoiding or responding to those counterexamples.

Note that none of this guarantees that such a philosopher will robustly know whether that position is true. After all, similarly situated philosophers in any given debate may disagree with her and will defend contrasting positions. These philosophers may agree, for instance, that certain counterexamples apply, but disagree about how damaging they

are. Or they may think that alternative positions cohere better with other intuitions or with other theoretical commitments. Even when all of these philosophers are acting in good faith and honestly attempting to discern the best reasons for particular positions, there will inevitably be some disagreement.

In order to avoid succumbing to the challenges of contingent selection and committee deference, a philosopher who accepts the picture I have painted here can adopt a role that relies on the contributions of philosophers as a community by conveying relevant philosophical arguments and inferences. In short, she can act as a *conduit for the discipline*, laying out the various potentially relevant normative positions and explaining how they would bear on the question at hand. She would *not* advocate for her own position (although perhaps she could mention which it is). Note that this role avoids contingent selection problems in two ways: (1) since she would not advocate her own view, but give it equal billing with others, it is less important or unimportant what her particular normative position is; and (2) since all philosophers attempting to give an overview of the current positions and their relevance would be mapping the same – or at least similar – normative terrain, there would be limited variation or contingent evaluation.

This would also avoid committee deference, because the goal is not to convince the committee of any substantive positions, but to convey the various normative arguments that support particular policy options. So even though she could potentially convince the committee if she were to try to argue for a particular position as correct, she holds her fire and refrains, allowing them to get a sense of the complexity of the philosophical debate. This avoids committee deference based on plausibility.

In order to avoid committee deference based on representation, the philosopher's evaluation relies on a key step: explicitly indicating where, and to what extent, there is normative argumentative convergence on specific policies. In particular, proposals that command support from different normative positions would be indicated as more robust, with greater robustness showing more philosophical justification for a particular policy. In this manner, in the place of simple 'morality (or justice) supports this policy' messages, there are more informative – and subtle – messages like 'this policy is supportable from a variety of normative positions' (or no positions, or a few positions), together with an explanation of which positions support that policy, which don't, and what the arguments are from those positions to those normative policy judgments. Note that, in contrast to the populist, the goal is not to 'count votes' or even to construct some generally unified picture – the normative positions in play may be in tension or even inconsistent – it is to consider the *range* of arguments.²⁸ So what is important is the relevance and validity of arguments, not the number of their adherents.

While the role I endorse is rarely explicitly defended as a specific position, several philosophers say things that are broadly indicative of it. For instance, Yarborough writes that 'There is broad consensus that if it is appropriate to present oneself as an expert, that expertise is revealed largely, some say exclusively, in presenting the various sides of an issue, citing relevant literature and summarizing the views to be found in that literature'.²⁹ Warnock says a key part of the philosopher's role on a policy committee is 'above all to consider the moral arguments on each side, such as they are, and to set them out with clarity'.³⁰ However, Brownlee makes claims that lie perhaps closest to the view here, asking whether 'weight should be given to any modest consensus amongst moral and political philosophers ... it sometimes may be appropriate to lean on the assumption that what ... is widely believed or agreed upon with that [normative] community should carry weight

in the public assessment of any one philosopher's practical recommendations'.³¹ I differ in some subtle ways, but one important one is that I believe it is not just important to convey when there is a (modest) consensus, but also when there is dissensus – and, most importantly, *why* in terms of the arguments and inferences underpinning these conclusions.

4.1. *Explicating the Convergent Evaluation Procedure*

With this motivation, it is worth providing a candidate procedure for what I call a 'convergent evaluation' for a philosopher playing the 'conduit role':

1. Identify the issue and current state of affairs
2. Determine the relevant arguments and normative positions espoused by philosophers*
3. Create a profile of possible solutions
4. Determine the level of convergence (and divergence) on those solutions generated by those arguments*
5. Evaluate policy options with reference to the level of convergence (and divergence) among philosophical positions*

Let me expand. In the case of carbon budgets, an example of a morally relevant issue is how to distribute emissions allowances (say, at a sectoral or regional level) given a national budget. With an identified morally relevant issue, we can consider an option that could be explicitly or implicitly adopted. An example of an option is that emissions should be distributed equally per capita. Another example could be that emissions should be distributed with sensitivity to the regional needs of private or public activities.

The second step requires sufficient understanding of the extant moral arguments and literature that one has a sense of the normative positions that have been developed and can evaluate the assumptions being made on that basis. Obviously, it is not necessary that one has comprehensive knowledge of the literature, but the more of the literature one is aware of, the closer one gets to being able to summarize it accurately. In this context, I would take standard first-order normative views (e.g. utilitarianism, Kantian deontology, Aristotelian virtue ethics) to all be distinct normative positions – but one might also take mid-level principles³² (e.g. the beneficiary pays principle or the polluter pays principle) or approaches to justice³³ (e.g. sufficientarianism, utilitarianism, or egalitarianism) to be distinct normative positions. Roughly speaking, I am taking the fact that a normative position or claim is defended in the (somewhat nebulous) mainstream published literature to make that position worthy of consideration. The process of summarizing will inevitably mean that some nuance and complexity will be lost, but the goal is to convey some of the influential or mainstream positions together with an indication of whether or to what extent arguments from those positions support various options. An important practical issue is that there may be trade-offs between comprehensiveness and accessibility; one may want to entice and engage the committee – but one should avoid going deep into the conceptual weeds and boring them (in the Appendix A, I indicate how I have tried to do this in practice).

Finally, in order to consider these policies, a convergent evaluation involves reporting levels of convergence. There are a variety of ways this could be done, but in the simple

system I used, I employed the following terms. When a key moral assumption could be supported by some philosophical positions, that assumption can be thought of as ‘philosophically plausible’, with more support from different positions indicating greater ‘philosophical robustness’. When it is outside of most or all of the range of positions, that assumption can be thought of as ‘philosophically objectionable’ or ‘questionable’. These terms seem to me concise and transparent, but one might prefer more fine-grained categories or different terminology. What is important is that those unacquainted with the philosophical literature or the field get a sense of how philosophers view the justification of the morally relevant assumptions being made. The resultant convergent evaluation is intended to be the *philosophical* contribution to the committee recommendation process; my claim is *not* that it is meant to be solely determinative.

4.2. *Why Adopt a Convergent Evaluation?*²

This subsection lays out a variety of justifications for adopting this convergent evaluator role in the policy committee context.

The first justification is to develop a methodology that draws on specifically philosophical skills and knowledge. As noted above, the populist’s evaluation seems to require expertise with empirically aggregating societal attitudes – not clearly a role for philosophers. The convergent evaluation draws on what I take to be the contribution that philosophers are best placed to make: the capacity to draw subtle, but valid, inferences from normative positions and relevant facts to practical conclusions. Normative philosophers are generally both better acquainted than non-philosophers with the relevant theories and especially well trained in carefully drawing valid inferences. At the very least, one can justify the convergent evaluation conditionally: supposing one believed philosophers do have some special competence to contribute, here is a candidate methodology that reflects that belief.³⁴ However, I would say something stronger. Individually, philosophers are trained to make valid inferences, so it is important what philosophers think about the link between normative positions and practical conclusions. With respect to a committee, philosophers are qualified to explicate and convey the various arguments and conclusions drawn by the philosophical community.

That leads to a second, more theoretical, justification: it is reasonable to assign a higher credence to an assumption supported by a variety of normative positions. This is for the familiar reason that, insofar as we are subject to normative or moral uncertainty, independent valid arguments for a conclusion make that conclusion more robust. That is not to say they *guarantee* truth; one might be able to support a moral conclusion with a variety of valid arguments, none of which are sound. But they do increase the likelihood of truth.³⁵ In that respect, reporting on convergence should increase the committee’s credence in the appropriate morally relevant assumptions – even if, as Wikler notes, these positions are often mutually incompatible.³⁶ More colloquially, Kamm points out that, in a committee report with a variety of normative reasons supporting the same conclusions, there is ‘a significant chance that among the various reasons in the report, a correct argument for the bottom line is present’.³⁷

A third justification for a convergent evaluation is purely pragmatic in the context of a committee. Members of a committee may well have different normative commitments or intuitions. Wolff writes that, on policy committees, ‘whether a moral view is correct, or right, or persuasive, takes second place to whether it is widely shared, or at least widely

accepted in the sense of enough people being prepared to live with it'.³⁸ Similarly, as Hirose writes of a (medical) policy committee, 'theoretical debate sometimes arises in ethics committees, divides the committee members, paralyzes the committee, and prevents the committee from arriving at a consensus about the ethical recommendations'.³⁹ I agree with Hirose that this practical challenge can be more easily addressed when a claim is supportable from more than one normative position. As a pragmatic matter, this helps motivate a convergent evaluation in spaces where policy recommendations are at issue among groups with heterogeneous normative views.

A fourth justification is procedural. Given that, on a committee, one plays the *role* of the philosopher (or ethicist or justice theorist), I believe that there are procedural justice reasons for that role to try to be reflective of philosophers. A convergent evaluation is a methodology for conveying a disciplinary conversation as opposed to a particular partisan's view. Similarly, I presume that a biodiversity expert on a committee is trying to convey what biologists and ecologists have discovered about biodiversity, not what *that* expert has found out about biodiversity. Furthermore, one might think that this role is procedurally important in the context of such an interdisciplinary committee: to some extent, one represents the discipline. This is true both in the sense that one should channel the views of the discipline, but also in the sense that this is a rare chance to expose others who may be unfamiliar with philosophy to a sense of the richness of the debates that occur within it (there is discussion of the value of such exposure).⁴⁰ In this manner, convergent evaluations also possess didactic virtues.

A fifth justification is that a convergent evaluation is intersubjectively verifiable, or at least more so than a partisan's evaluation. Suppose I say that a particular carbon budget is appropriate because it satisfies some kind of sufficiency threshold of distributive justice that I endorse. If a partisan who adopts a different shape of justice – for example, a prioritarian one – is considering my evaluation, they will likely judge my recommendations and my conclusions incorrect (supposing that they think that, from a prioritarian perspective, the recommended carbon budget is inferior to another). Contrast this with convergent evaluators. Suppose I report that a particular carbon budget is distributionally just from a sufficientarian normative position (and, say, also from an egalitarian one). A prioritarian can agree that the budget is just *from those normative positions*, while disagreeing, as a first-order normative matter, about whether sufficientarianism, egalitarianism, or prioritarianism is the appropriate shape of justice. In short, first-order intersubjective normative agreement is unnecessary for intersubjective verifiability, which convergent evaluation allows for.

A sixth justification comes from political philosophy and political theory. There are a variety of ways of expressing the idea, but the Rawlsian pursuit of 'overlapping consensus' is a relevant precedent.⁴¹ For Rawls, different citizens might have different 'comprehensive normative doctrines' with different associated theories of justice. But they can still agree on basic principles for society, as long as those basic principles can be supported with reference to a variety of doctrines. The claim I would make is similar, but perhaps more limited: policies that can be justified from a variety of normative perspectives are both likely to be viewed as, and plausibly be, more politically legitimate, than those that are only supportable from one.⁴² Note that to the extent that convergent evaluation confers greater acceptability, unlike with Wolff's or Warnock's suggestions, it does *not* do so by reflecting or explicitly considering public or social sentiment. In short, I believe that the convergence in a convergent evaluation can reflect a greater diversity of values for

those considering the evaluation – although there is no guarantee that it will reflect the values of any given reader! – and it is a procedure that could lend legitimacy to these recommendations in procedural justice terms.

Finally, while all of these justifications have different normative starting points and commitments, they all converge on supporting a convergent evaluation compared to a partisan's evaluation. Convergent evaluation says that, when a particular normative conclusion is supportable from a variety of normative positions, that makes it philosophically robust. So it is conceptually satisfying that a convergent evaluation of these different normative justifications itself helps justify convergent evaluations. In other words, there is metatheoretical self-recommendation for the methodology, or we could say that the methodology can be seen as metaconsistent. I take this to also recommend the methodology.

5. An Objection: What If Philosophers' Positions Are Radical?

An objection that could be raised to a convergent evaluation appeals to the (sometimes) radical views that philosophers have. For instance, in environmental ethics, there is a very strong consensus from a variety of normative starting points that meat-eating is very difficult to defend and perhaps morally impermissible.⁴³ In climate ethics, regardless of substantive normative positions, there is a strong consensus that morality and justice demand much more stringent policies than are observed in most wealthy countries. These positions may well be outside the mainstream, both socially and politically. Would I defend convergent evaluation procedures even in cases where philosophers converge on radical views?

The answer is yes, but let me split my responses between practical and theoretical considerations.

First, practically speaking, a committee is often given some set of issues and some policy options to consider. In such cases, the fact that philosophers would endorse outcomes that are stronger than the available policy options is not objectionable. Among the constrained set of options, the one that is closest to the philosophically most justifiable one could be considered subject to the strongest philosophical support.⁴⁴ That even stronger policies would be philosophically recommendable is neither concerning nor practically relevant.

Second, even if philosophers held a radical view – suppose that the vast majority of normative views converge on the claim that eating meat is morally impermissible – that does not straightforwardly imply that policy coercing citizens to avoid a morally impermissible action is morally obligatory, or even permissible. For instance, although lying is morally wrong, that does not imply that states can morally enforce truth-telling.⁴⁵ We might think that there are moral values (autonomy and liberty) that tell against instructing, let alone forcing, the public to never engage in vicious behaviour.

Third, theoretically speaking, it strikes me as entirely appropriate that a philosophical expert could convey a radical view on the basis of philosophical diversity of arguments supporting that view. But it is important to remember that a (joint) committee policy should reflect a variety of considerations – what is philosophically justifiable according to the (individual's) philosophical evaluation need not be dispositive. That is why I noted the distinction between the individual's evaluation and the committee recommendation.

Consider a potential analogy. A biodiversity expert might be called on to provide the scientific perspective of her field on whether to build some particular energy source in a

city. She might say that refraining from building it would have very positive benefits for biodiversity. But from the purely scientific point of view, perhaps tearing down the whole city to let wildlife develop would be the best outcome for biodiversity. Of course this is not something that is going to be done or recommended, but it is compatible with biodiversity being the value she is there to represent. This does not seem objectionable, even if her role would have more radical implications if all actions were on the table. After all, her individual evaluation is not an all-things-considered recommendation, nor should it necessarily be.

Fourth, theoretically speaking, I think this is not a bug, but a feature. I believe it extremely important that the methodology does not always or necessarily conform to societal or political expectations but is capable of delivering radical verdicts. Individual evaluation verdicts are subject to discussion and compromise, but if their independence is relinquished *before* evaluation, then that taints the evaluation.

Finally, trying to frame the evaluation in ways that will be acceptable to committee members and the policy process is not the special competence of the philosopher. If the goal is developing an evaluation that would convince others and pass through the policy process, then a social psychologist or a political scientist could be better placed than a philosopher.⁴⁶

More importantly, if the philosopher is trying to include various questions of feasibility before she even begins her evaluation, then her evaluation is subject to the challenges noted above. First, the specific way that she adjusts her evaluation in order to conform to others' views will represent judgment calls that are likely to vary greatly depending on who is appointed, making a version of contingent selection relevant here. Second, and potentially even more troubling, the value judgments about how to adjust her evaluation will be purposefully opaque to the committee; they will not know which part of her evaluation is what she thinks is philosophically justifiable and which part is adjusted to be more acceptable to others. This risks making committee deference even more of an issue, since it could be unclear which parts of the philosopher's evaluation are generated by philosophical considerations and which are generated by attempts to make those judgments politically amenable. If there is the risk that a committee will take views that are aimed at political acceptability to be required by morality or justice, that seems to me seriously procedurally objectionable.

6. Conclusion

Increasingly, various versions of public philosophy⁴⁷ are being developed and experimented with – but these advances also require more systematization about how public philosophy is to be done. I would suggest that this is an especially important issue in instances where social legitimacy and procedural justice are demanded – perhaps the apex of which involves public policy.

This article is an attempt to explicate a role ('the convergent evaluator') and contrast it with what I take to be the two main alternatives ('the partisan' and 'the populist'). It does so by drawing attention to two challenges that threaten to arise in the context of policy committees, challenges that have not been discussed in the philosophical literature.

This is also meant to instigate discussion about how philosophers can contribute to public life, but in a more methodologically explicit and sophisticated manner.

Non-philosophers have little idea about what philosophers do or how we come to our positions (as opposed to empirical disciplines which usually appeal to data and some other humanities disciplines where the goal is to deeply understand a text or historical event). It is therefore worthwhile to be able to present or at least point to explicit methodologies and procedures to answer these concerns. I hope that developing explicit methodologies allows philosophers to contribute to public policy debates more confidently and constructively.

Kian Mintz-Woo, Philosophy and Sustainability Institute, University College Cork, Cork, Ireland; Equity and Justice Group, International Institute for Applied Systems Analysis, Laxenburg, Austria. mintzwoo@ucc.ie

Acknowledgements

Part of the discussion draws on voluntary committee work that the author engaged in for the Irish government's Carbon Budget Working Group (2024–5). Thanks to Pierre André, Alyssa R Bernstein, Megan Blomfield, Vittorio Bufacchi, Blake Francis, Meabh Gallagher, Francisco Garcia-Gibson, Laura García-Portela, Ewan Kingston, Philip O'Brien, Don Ross, and Diana Piroli as well as others in the Carbon Budget Working Group for thoughts and input. Thanks to audiences at the University of Oxford (especially Hayden Wilkinson, Jo Wolff, Joel Martinsson, Johan Go, Jonathan Floyd, and Doug Bamford), Harvard University (especially Britta Clark, Arthur Obst, Ronald Sandler, Lucas Stanczyk, and Kathrin von Allmen), and Sant'Anna School of Advanced Studies (especially Diana Piroli and Alberto Pirni). Many thanks to the editors, including the guest editors, as well as to the anonymous reviewers, for a variety of helpful comments. Open access funding was provided by IReL, the Irish library consortium.

NOTES

- 1 It is worth noting that carbon budgeting, i.e. considering only the quantity of carbon emissions as opposed to their timing, is sensible because of some chemical quirks for carbon dioxide. Most pollutants, including other greenhouse gases like methane, are much more sensitive to timing of emissions than total quantity. The ethics of such 'flow' pollutants (like methane), as opposed to 'stock' pollutants (like carbon dioxide), is relatively unexplored in the climate ethics literature and deserves careful consideration.
- 2 There is the further question of whether carbon budgets are justified at all, i.e. whether they are the appropriate policy tool for structuring or planning mitigation; Tank, "Against the Budget." I would set this question aside as beyond the scope of this article, but it is worth making a few quick points in defence of carbon budgets. The main, negative, point is that a highly differentiated approach based on subsistence and luxury emissions as Tank envisions requires too much information to be practically applicable in a policy context. Not only would it obviously require thresholds for subsistence and luxury emissions (which would be even more complex and politically contentious than carbon budgets already are), but it would require national or international agreement on which countries are beyond subsistence and luxury limits. Another, more positive, point is that even supposing we did find such thresholds, we would still be engaged in carbon budgets; we would just have different budgets that were associated with different goals (subsistence and luxury). This would not undermine carbon budgets, but simply generate a more complex and fine-grained set of budgets.
- 3 Note that 'position' is to be construed very widely and, importantly, to include edge cases like moral particularism, whereby there are no principles or general rules (virtue ethics is sometimes also taken to be

anti-theoretical in this sense, but it counts as a position on this wide reading). It is meant to be a neutral term which encompasses a range of normative commitments.

- 4 Naturally, the roles discussed here are ideal types and not necessarily exclusive; in actual circumstances, hybrids of these roles could be constructed. However, I think this is a reasonable simplification since it seems likely that the philosopher will take some role as especially salient and would therefore conduct her methodological practice accordingly. Additionally, 'populist' is sometimes used to refer to some thicker (and perhaps more objectionable) political views. The name here merely reflects the adoption of political judgments with reference to public views and is not meant to indicate these thicker political uses of the term.
- 5 Perhaps this role is not completely disjoint from a philosopher's general working duties. In particular, I suspect that the role reflected in the philosophers' evaluation is quite similar to the role some of us think we should have when teaching; namely, to give a fair hearing to a variety of positions but not to proselytize our own philosophical positions. In my view, this analysis suggests a clear reason: the partisan and her attendant adversarial role are appropriate when there are a variety of similarly expert interlocutors who are mutually engaged in truth-seeking. To put the point somewhat polemically, my position is that truth-seeking is not the direct purpose of a lecture or a policy committee.
- 6 Cook *et al.*, "Consensus."
- 7 Bourget and Chalmers, "What"; Bourget and Chalmers, "Philosophers."
- 8 Kamm, "Some Points," 147, expresses a similar concern.
- 9 Key examples discussed by Archard, "Mary Warnock"; Brody, "Introduction"; Hirose, *Ethics of Pandemics*; Toulmin, "Tyranny"; Weisbard, "Role"; Weisbard and Arras, "Commissioning Morality"; Wikler, "Introduction."
- 10 Archard, "Philosophical Advice"; Brock, "Truth"; Kamm, "Philosopher"; Warnock, "Moral Thinking"; Wikler, "Bioethics"; Wolff, "Method"; Wolff, *Public Policy*.
- 11 Wolff, "Method."
- 12 Brock, "Truth", 790 (emphasis added); Wikler, "Introduction."
- 13 Momeyer, "Philosophers," 398.
- 14 Kamm, "Philosopher," 357.
- 15 *Ibid.*, 358.
- 16 Wolff, "Method," 17–21.
- 17 Fletcher, "Bioethics," 317.
- 18 Kamm, "Philosopher," 357.
- 19 It is possible that the populist's evaluation appeals to some suitably informed or systematized (hypothetical) version of societal values; cf. Archard, "Moral Philosophers." One potential way of getting evidence about what members of society would decide were they to be suitably informed is to take a representative sample, inform them and let them deliberate, and see what they conclude. This is the goal of the type of mini-public we sometimes call 'citizens' assemblies' and it has been argued that they provide normatively relevant evidence about how citizens in general would decide, were they to be similarly informed and allowed to deliberate together; cf. Mintz-Woo, "Carbon Pricing." I take this to be a version of the populist's appeal.
- 20 Archard, "Mary Warnock," 754.
- 21 Also cf. Miller, "Distributive Justice"; Walzer, *Spheres*.
- 22 Wolff, "Method."
- 23 This dilemma is structurally similar to the debate in American legal systems about whether to interpret the Constitution in terms of its evolving use and application or to attempt to limit interpretation to strict intention of original authors. My concerns about philosophers not being trained social scientists is not dissimilar to the concern that judges are not trained historians.
- 24 O'Neill, "Applied Ethics," 220.
- 25 Kamm, "Philosopher," 373.
- 26 It is open for defenders of the populist role to reply that philosophical skills in systematizing or surfacing deep values outweigh the importance of a social scientist's skills in empirically understanding societal values. While I am sceptical of this claim, even if it were granted *arguendo* I believe objectionable conventionalism is still a decisive response to adopting the populist role.
- 27 Hannon and Nguyen, "Understanding."
- 28 Thanks to a reviewer for flagging the similarity of this methodology to the ethical matrix methodology; cf. Mepham, "Ethical Analysis"; Mepham, "Framework." Ethical matrices have rows of stakeholders and columns of values or normative positions, trying to systematically determine how different choices affect

stakeholders. While this methodology likely has similar motivations to convergent evaluation, there are some important differences.

First, on the input side, an ethical matrix presumes a particular set of values. Mepham, “Framework,” takes Beauchamp and Childress’s (*Principles*) principlism as a neutral source of values. Obviously, there are a variety of first-order reasons to object, whether because one takes a different moral theory to be more appropriately neutral (Greene, *Moral Tribes*) or because one finds principlism indeterminate or underspecified (Wikler, “Bioethics,” 239). As a methodological criticism, however, I would point out that the relevant set of values is underdetermined by the methodology. In contrast, the role of a disciplinary conduit supports the appeal to normative positions taken to be philosophically plausible for the convergent evaluator.

Second, on the output side, I believe the ethical matrix methodology has two challenges. If the cells do not encode substantive philosophical or normative analysis (cf. ‘the Matrix is in principle ethically neutral, i.e., it is an *analytical* tool’; Mepham, “Framework,” 169, emphasis his), then simply adopting the value does not suffice to know what to do. That is because the values may conflict. Hence, someone who adopts the conflicting values just knows that and for whom they conflict. In contrast, the convergent evaluation explains what normative positions under consideration imply for the policy question at hand – if you adopt the normative position, the evaluation provides practical judgments.

Most importantly, insofar as the cells give information about morally relevant implications, they do not make it transparent *how* outcomes relate to the given values. Assuming that there is some substantive philosophical work that goes into linking these values to the implications (if not, it does not seem helpful to have philosophers filling in the ethical matrices), the reasoning is suppressed in the matrix. In contrast, convergent evaluation surfaces and explicates the relevant moral reasoning. For these reasons, I find convergent evaluation more justifiable than the ethical matrix methodology.

29 Yarborough, “Reluctant,” 348–9.

30 Warnock, “Moral Thinking,” 513.

31 Brownlee, “Normative Principles,” 232–3.

32 Mintz-Woo, “Compensation Duties.”

33 Zimm *et al.*, “Justice Considerations.”

34 Note that this methodology supports the view that philosophers have special competences – without relying on the controversial claim that expertise with moral theory implies expertise in moral judgment *simpliciter*.

35 Technically speaking, there are a couple of background assumptions here: I am assuming that the variety of normative positions at issue are all potentially true (i.e. have greater than zero probability) and that they are probabilistically independent (i.e. if utilitarianism supports a conclusion, consequentialism’s supporting the same conclusion should not increase one’s credence in that conclusion since utilitarianism and consequentialism are not probabilistically independent).

36 Wikler, “Bioethics,” 245.

37 Kamm, “Philosopher,” 352. Also cf. Toulmin, “Tyranny.”

38 Wolff, *Public Policy*, 4.

39 Hirose, *Ethics of Pandemics*, 9.

40 Olasov, “Public.”

41 Rawls, *Theory of Justice*.

42 Of course, while I believe that this is generally true, it need not be universally true. Consider a case where a supermajority has some particular normative position and the rest of the people have a variety of normative positions, but the latter positions all converge on a particular policy. While this policy might be supportable from a variety of normative positions, it is still only supported by a (small) minority.

43 DeGrazia, “Moral Philosophy,” 50, writes concisely that rejecting anthropocentrism is a ‘modest moral claim ... yet full implementation ... would change the world’.

44 Benjamin, “Philosophical Integrity.”

45 Momeyer, “Philosophers.”

46 In support of my point, it is worth noting that Kamm, “Philosopher,” 371, reports that Brock’s (“Truth”) second-guessing of the consequences of his corrections to committee colleagues was mistaken; even after Brock explained what he took to be flaws in their views, they retained their commitment to their original conclusions. This lends credence to my position: philosophers are not experts with respect to the political consequences of their argumentation, at least in committee contexts.

47 McIntyre *et al.*, *Companion*.

48 Mintz-Woo, *Irish Carbon Budgets*.

49 Dooley *et al.*, “Ethical Choices”; Hirose, *Ethics of Pandemics*; Zimm *et al.*, “Justice Considerations.”

- 50 Zimm *et al.*, “Justice Considerations.”
- 51 Mintz-Woo, “Compensation Duties.”
- 52 Fleurbaey *et al.*, “Sustainable Development.”
- 53 The policy principle of ‘contraction and convergence’, whereby different countries evolve their emissions closer to some roughly equal per capita emissions, is a way of implementing a version of global emissions equality.
- 54 This involves some simplification – since there are distributional justice approaches to intergenerational and intertemporal provision – but I think the simplification is justified by the utility of a conceptual map for those who are unfamiliar with these debates.
- 55 Gardiner, “Climate Targets.”
- 56 Caney, “Struggle”, 9f; Caney, “Governance”; Jamieson, *Reason*, 227f; McKinnon, *Future Justice*, 4f; Moellendorf, *Moral Challenge*, 23ff.
- 57 Caney, “Human Rights,” 538.
- 58 Steel *et al.*, “Collapse.”
- 59 Also cf. McKinnon, *Future Justice*; Steel *et al.*, “Civilization.”
- 60 Among the few exceptions, two stand out. First, Bovens, “Lockean Defense,” argues that if emissions are such that ‘enough and as good’ (emissions capacities) are left for others (the ‘Lockean proviso’), then it is not a problem for some to profitably or productively emit. However, the existing, and plausibly historical, carbon budgets are sufficiently limited that the Lockean proviso suggested is not met. When some are emitting in our current circumstances, that does *not* leave comfortable space in carbon budgets for others. Second, Meyer and Sanklecha, “Legitimate Expectations,” argue that when governments have policies that generate legitimately expectations on behalf of citizens (i.e. that those policies will continue), the fact that citizens build their lives around those expectations can be enough to justify strong constraints on changing these policies (also cf. Meyer and Sanklecha, “Individual Expectations”). This ‘legitimate expectations’ theory could be read as justifying continuing policies that facilitate current (high) emissions. However, as Culp, “Comment,” points out, it is not clear that current emissions are compatible with justice, especially with intergenerational justice, so they may not get the initial benefit of the doubt. Regardless, Meyer and Sanklecha do not explicitly defend grandfathering on the basis of this legitimate expectations theory.
- 61 Schulan *et al.*, “Distributive Justice,” 6.
- 62 Moellendorf, *Moral Challenge*.
- 63 While many moral philosophers, especially utilitarians, think we should include non-human sentient animals in the scope of justice, this is almost unheard of in climate policy. This may be another, small, way in which Ireland’s assumptions are favourable to Ireland: the harms to non-humans from climate change may be orders of magnitude of the harms to humans, such that a justifiable carbon budget would be more limited than if only anthropocentric harms were considered.
- 64 Singer, *One World*.

References

- Archard, David. “Why Moral Philosophers Are Not and Should Not Be Moral Experts.” *Bioethics* 25, no. 3 (2011): 119–127. <https://doi.org/10.1111/j.1467-8519.2009.01748.x>.
- Archard, David. “Philosophical Advice.” *Philosophy* 96, no. 4 (2021): 603–623. <https://doi.org/10.1017/S0031819121000139>.
- Archard, David. “Mary Warnock and ‘Public Philosophy’.” *Journal of Applied Philosophy* 41, no. 5 (2024): 751–766. <https://doi.org/10.1111/japp.12747>.
- Beauchamp, Tom, and James F. Childress. *Principles of Biomedical Ethics*, 4th ed. New York: Oxford University Press, 1994.
- Benjamin, Martin. “Philosophical Integrity and Policy Development in Bioethics.” *Journal of Medicine and Philosophy* 15, no. 4 (1990): 375–389. <https://doi.org/10.1093/jmp/15.4.375>.
- Bourget, David, and David J. Chalmers. “What Do Philosophers Believe?” *Philosophical Studies* 170, no. 3 (2014): 465–500. <https://doi.org/10.1007/s11098-013-0259-7>.
- Bourget, David, and David J. Chalmers. “Philosophers on Philosophy: The 2020 PhilPapers Survey.” *Philosophers’ Imprint* 23 (2023): 11. <https://doi.org/10.3998/phimp.2109>.

- Bovens, Luc. "A Lockean Defense of Grandfathering Emission Rights." In *The Ethics of Global Climate Change*, edited by Denis G. Arnold, 124–144. Cambridge: Cambridge University Press, 2011.
- Brock, Dan W. "Truth or Consequences: The Role of Philosophers in Policy-Making." *Ethics* 97, no. 4 (1987): 786–791.
- Brody, Baruch A. "Introduction." *Journal of Medicine and Philosophy* 15, no. 4 (1990): 345–46. <https://doi.org/10.1093/jmp/15.4.345>.
- Brownlee, Kimberley. "Normative Principles and Practical Ethics: A Response to O'Neill." *Journal of Applied Philosophy* 26, no. 3 (2009): 231–37. <https://doi.org/10.1111/j.1468-5930.2009.00447.x>.
- Caney, Simon. "Human Rights, Climate Change, and Discounting." *Environmental Politics* 17, no. 4 (2008): 536–555. <https://doi.org/10.1080/09644010802193401>
- Caney, Simon. "The Struggle for Climate Justice in a Non-Ideal World." *Midwest Studies in Philosophy* 40, no. 1 (2016): 9–26. <https://doi.org/10.1111/misp.12044>.
- Caney, Simon. "Global Climate Governance, Short-Termism, and the Vulnerability of Future Generations." *Ethics & International Affairs* 36, no. 2 (2022): 137–155. <https://doi.org/10.1017/S0892679422000181>.
- Cook, John, Naomi Oreskes, Peter T. Doran, William R. L. Anderegg, Bart Verheggen, Ed W. Maibach, et al. "Consensus on Consensus: A Synthesis of Consensus Estimates on Human-Caused Global Warming." *Environmental Research Letters* 11 (2016): 048002. <https://doi.org/10.1088/1748-9326/11/4/048002>.
- Culp, Julian. "Comment on Lukas Meyer and Pranay Sanklecha. Individual Expectations and Climate Justice." *Analyse & Kritik* 33, no. 2 (2011): 473–76. <https://doi.org/10.1515/auk-2011-0209>.
- DeGrazia, David. "Does Moral Philosophy Have Any Point?" In *What is the Point of Moral Philosophy*, edited by Roger Crisp, Ingmar Persson, and Julian Savulescu, 46–62. Abingdon: Routledge, 2026. <https://doi.org/10.4324/9781003496090-4>.
- Dooley, Kate, Ceecee Holz, Sonja Klinsky, Sivan Kartha, J. Timmons Roberts, Henry Shue, Harald Winkler, et al. "Ethical Choices Behind Quantifications of Fair Contributions Under the Paris Agreement." *Nature Climate Change* 11, no. 4 (2011): 300–305. <https://doi.org/10.1038/s41558-021-01015-8>.
- Fletcher, John C. "Bioethics in a Legal Forum: Confessions of an 'Expert' Witness." *Journal of Medicine and Philosophy* 22, no. 4 (1997): 297–344. <https://doi.org/10.1093/jmp/22.4.297>.
- Fleurbay, Marc, S. Kartha, S. Bolwig, Y. L. Chee, Y. Chen, E. Corbera, et al. "Sustainable Development and Equity." In *Climate Change 2014: Mitigation of Climate Change: Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Ottmar Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, et al., 283–350. New York: Cambridge University Press, 2014.
- Gardiner, Stephen M. *Climate Targets and Moral Corruption*. Newark, Delaware: American Philosophical Association Blog, 2023. <https://blog.apaonline.org/2023/04/10/climate-targets-and-moral-corruption/>.
- Greene, Joshua. *Moral Tribes*. New York: Penguin, 2014.
- Hannon, Michael, and James Nguyen. "Understanding Philosophy." *Inquiry* (forthcoming). <https://doi.org/10.1080/0020174X.2022.2146186>.
- Hirose, Iwao. *Ethics of Pandemics*. Abingdon: Routledge, 2023. <https://doi.org/10.4324/9781003203759>.
- Jamieson, Dale. *Reason in a Dark Time*. Oxford: Oxford University Press, 2014.
- Kamm, Frances M. "The Philosopher as Insider and Outsider." *Journal of Medicine and Philosophy* 15, no. 4 (1990): 347–374. <https://doi.org/10.1093/jmp/15.4.347>.
- Kamm, Frances M. "Some Points of Moral Philosophy and Obstacles They May Create." In *What is the Point of Moral Philosophy*, edited by Roger Crisp, Ingmar Persson, and Julian

- Savulescu, 138–152. Abingdon: Routledge, 2026. <https://doi.org/10.4324/9781003496090-11>.
- McIntyre, Lee C., Nancy Arden McHugh, and Ian Olosov, eds. *A Companion to Public Philosophy*. Hoboken: Wiley, 2022.
- McKinnon, Catriona. *Climate Change and Future Justice: Precaution, Compensation, and Triage*. Abingdon: Routledge, 2012. <https://doi.org/10.4324/9780203802205>.
- Mephram, Ben. "Ethical Analysis of Food Biotechnologies: An Evaluative Framework." In *Progress in Dairy Science*, edited by Ben Mephram, 375–395. Wallingford: CAB, 1996.
- Mephram, Ben. "A Framework for the Ethical Analysis of Novel Foods: The Ethical Matrix." *Journal of Agricultural and Environmental Ethics* 12, no. 2 (2000): 165–176. <https://doi.org/10.1023/A:1009542714497>.
- Meyer, Lukas H., and Pranay Sanklecha. "Individual Expectations and Climate Justice." *Analyse & Kritik* 33, no. 2 (2011): 449–471. <https://doi.org/10.1515/auk-2011-0208>.
- Meyer, Lukas H., and Pranay Sanklecha. "How Legitimate Expectations Matter in Climate Justice." *Politics, Philosophy & Economics* 13, no. 4 (2014): 369–393. <https://doi.org/10.1177/1470594X14541522>.
- Miller, David. "Distributive Justice: What the People Think." *Ethics* 102, no. 3 (1992): 555–593. <https://doi.org/10.1086/293425>.
- Mintz-Woo, Kian. "Compensation Duties." In *Handbook of Philosophy of Climate Change*, edited by Gianfranco Pellegrino and Marcello di Paola, 779–797. Cham: Springer, 2023.
- Mintz-Woo, Kian. "Carbon Pricing Is Not Unjust." *Global Challenges* 8, no. 1 (2024a): 2300089. <https://doi.org/10.1002/gch2.202300089>.
- Mintz-Woo, Kian. *Irish Carbon Budgets: Some Moral Considerations*. Dublin: Climate Council, 2024b. <https://www.climatecouncil.ie/carbonbudgets/carbonbudgetsworkinggroup2023-2024/>.
- Moellendorf, Darrel. *The Moral Challenge of Dangerous Climate Change: Values, Poverty, and Policy*. Cambridge: Cambridge University Press, 2014.
- Momeyer, Richard W. "Philosophers and the Public Policy Process: Inside, Outside, or Nowhere at All." *Journal of Medicine and Philosophy* 15, no. 4 (1990): 391–409. <https://doi.org/10.1093/jmp/15.4.391>.
- O'Neill, Onora. "Applied Ethics: Naturalism, Normativity and Public Policy." *Journal of Applied Philosophy* 26, no. 3 (2009): 219–230. <https://doi.org/10.1111/j.1468-5930.2009.00446.x>.
- Olosov, Ian. "Public." *Philosopher* 110, no. 4 (2022). <https://www.thephilosopher1923.org/post/the-new-basics-public>.
- Rawls, John. *A Theory of Justice*. Belknap: Cambridge, MA, 1971.
- Schulan, Alexander, Lukas Tank, and Christian Baatz. "Distributive Justice and the Global Emissions Budget." *WIREs Climate Change* 14, no. 5 (2014): e847. <https://doi.org/10.1002/wcc.847>.
- Singer, Peter. *One World: The Ethics of Globalisation*. New Haven: Yale University Press, 2002.
- Steel, Daniel, C. Tyler DesRoches, and Kian Mintz-Woo. "Climate Change and the Threat to Civilization." *Proceedings of the National Academy of Sciences* 119, no. 42 (2022): e2210525119. <https://doi.org/10.1073/pnas.2210525119>.
- Steel, Daniel, Kian Mintz-Woo, and C. Tyler DesRoches. "Collapse, Social Tipping Dynamics, and Framing Climate Change." *Politics, Philosophy & Economics* 23, no. 3 (2024): 230–251. <https://doi.org/10.1177/1470594X231196432>.
- Tank, Lukas. "Against the Budget View in Climate Ethics." *Critical Review of International Social and Political Philosophy* 27, no. 7 (2024): 1075–88. <https://doi.org/10.1080/13698230.2022.2070833>.
- Toulmin, Stephen. "The Tyranny of Principles." *Hastings Center Report* 11, no. 6 (1981): 31–39. <https://doi.org/10.2307/3560542>.

- Walzer, Michael. *Spheres of Justice: A Defense of Pluralism and Equality*. New York: Basic Books, 1983.
- Warnock, Mary. "Moral Thinking and Government Policy: The Warnock Committee on Human Embryology." *Milbank Memorial Fund Quarterly* 63, no. 3 (1985): 504–522. <https://doi.org/10.2307/3349845>.
- Weisbard, Alan J. "The Role of Philosophers in the Public Policy Process: A View from the President's Commission." *Ethics* 97, no. 4 (1987): 776–785. <https://doi.org/10.1086/292890>.
- Weisbard, Alan J., and John D. Arras. "Commissioning Morality: An Introduction to the Symposium." *Cardozo Law Review* 6, no. 2 (1984): 223–241.
- Wikler, Daniel. "Introduction." *Ethics* 97, no. 4 (1987): 775. <https://doi.org/10.1086/292889>.
- Wikler, Daniel. "What Has Bioethics to Offer Health Policy?" *Milbank Memorial Fund Quarterly* 69, no. 2 (1991): 233–251. <https://doi.org/10.2307/3350203>.
- Wolff, Jonathan. "Method in Philosophy and Public Policy: Applied Philosophy Versus Engaged." In *The Routledge Handbook of Ethics and Public Policy*, edited by Annabelle Lever and Andrei Poama. Abingdon: Routledge, 2019.
- Wolff, Jonathan. *Ethics and Public Policy: A Philosophical Inquiry*, 2nd ed. Abingdon: Routledge, 2020.
- Yarborough, Mark. "The Reluctant Retained Witness: Alleged Sexual Misconduct in the Doctor/Patient Relationship." *Journal of Medicine & Philosophy* 22, no. 4 (1997): 345–364. <https://doi.org/10.1093/jmp/22.4.345>.
- Zimm, Caroline, Kian Mintz-Woo, Elina Brutschin, Susanne Hanger-Kopp, Roman Hoffmann, Jarmo S. Kikstra, et al. "Justice Considerations in Climate Research." *Nature Climate Change* 14, no. 4 (2024): 22–30. <https://doi.org/10.1038/s41558-023-01869-0>.

Appendix A

Convergent Evaluation Applied to Carbon Budgets

Having provided an explanation of the convergent evaluation procedure and some motivations for adopting it, in this Appendix A, I turn to applying it in the case of carbon budgets. This section can be seen as both helping illustrate the methodology and as a case study of how it has been used in an actual policy process, drawing from the author's experiences working on Irish carbon budget methodology as part of an expert working group (for context, the group was around a dozen people, the author was the only philosopher, and various parts of civil society including thinktanks were represented alongside a handful of academics). I begin with the issues (Section A.1) and continue with evaluating with respect to convergence (Section A.2).

In the normative evaluation I performed for a carbon budget policy committee, we had a simple model for burden-sharing. The model determined, based on, *inter alia*, the Paris Agreement limits, what the limits would be for Ireland. Identifying the key morally relevant values that underpinned these policy options required moderately sophisticated understanding of the simple model together with several discussions with relevant stakeholders and experts.

A.1. *Identifying the Issues*

How does one determine which issues are morally relevant and drive the evaluation of policy options? Here, there is no algorithmic answer; one needs to be both sufficiently familiar with the models or formulae involved and their moral salience to make a judgment. I also think it worth iterating on initial sets of issues, in the sense that presenting, tentatively, some set of issues one takes to be morally relevant can elicit helpful feedback.

Once a set of morally relevant issues have been identified, another problem may arise. In any given policy context, it may well turn out that some issues have not been discussed or recognized in the philosophical literature. This may be because philosophers think the answer obvious or because the issue has been overlooked. These should, of course, be explicitly noted, but it may be inaccurate to say they are philosophically robust or that they are philosophically non-robust.

Of the issues that are both morally relevant and discussed by philosophers, these can be the most useful and informative places to engage in a convergent evaluation. For some philosophers, it might be necessary to check whether, or how much, philosophers have discussed the assumptions in question – but ideally a philosopher in the relevant role will already have a good general understanding of the relevant debates and literatures.

For those interested specifically in the ethics of carbon budgets, I will explain the morally relevant issues I identified in the carbon budget policy committee I took part in.⁴⁸

The first morally relevant issue was about the (global) temperature limit. This will generally be a key initial assumption in carbon budgets since it sets a global constraint on what is allowed. Sometimes, the limit will be strict (it cannot be breached); more often, some overshoot is allowed with some form of sinks reducing the emissions in the system (sinks can be artificial or natural). Besides the limit, the choice of whether the limit is strict or is the long-run equilibrium or the temperature by a specific date is also a morally relevant issue.

The second morally relevant issue was the carbon budget reference year. Relative to what baseline year is a carbon budget to be measured? Beyond the obvious moral questions (which or how many historical emissions are grandfathered in), there are also some relevant practical issues involved: which years have the most robust data? Are there exogenous or endogenous shocks that might make some years highly unrepresentative? Are there other peer countries with similar reference years, making for transparent comparisons? However, one less obvious point is that a more stringent target and a later (higher emission) reference year might give the same carbon budget as a less stringent target and an earlier (lower emission) reference year, so in at least one important respect this can be seen as an accounting assumption.

The third, and perhaps most vexed, morally relevant issue was which ‘equity principle’ to adopt. This requires some translation: in policy contexts, terms like ‘equity principles’ and ‘fairness’ are often used, but they are rather imprecise and subjective compared to the way philosophers usually approach distributional questions.⁴⁹ In this part of the report, I tried to explain to non-philosophers some of the important debates in this space. More specifically, I distinguished between what I call ‘justice-focused climate ethicists’ and ‘principle-focused climate ethicists’. This is not a sharp distinction, but I thought it didactically prudent to roughly divide up two debates which often happen independently. On the justice side, we have theories of distributional justice with (typically) a scope of justice, a metric (or currency) of justice, and a shape (or pattern) of justice.⁵⁰ I explained that the

scope indicates who the theory applies to; the metric is the (morally relevant) quantity subject to the distribution; and the pattern indicates how the metric is ideally distributed. To give examples, I indicated that the scope could be national or cosmopolitan, the metric could be consumption or welfare, and the shape could be egalitarian, sufficientarian, utilitarian, or prioritarian. On the principles side, we have mid-level principles that have emerged from the climate ethics literature, with the dominant three (not necessarily exclusively) being the polluter pays principle (PPP), the beneficiary pays principle (BPP), and the ability to pay principle (APP).⁵¹

After introducing these two debates, I noted that the equity principles that have arisen (especially in the Intergovernmental Panel on Climate Change's Fifth Assessment Report, AR5) in the climate policy context map imprecisely on parts of these debates. The equity principles in the third working group of the IPCC AR5⁵² are: responsibility, capacity or ability to pay, equality, and a right to development.⁵³ Responsibility is close to PPP and capacity is a version of APP, but equality could refer to any of a set of distributional justice theories (albeit without a specified metric of justice) and a right to development is harder to pinpoint, but I believe is best justified by some type of sufficientarian theory of distributional justice.

However, looking at the two climate ethics debates (justice-focused and principle-focused) yields a larger space of normative positions as well as clarity on the ways that juxtaposing an 'equality' equity principle with a 'responsibility' equity principle is somewhat awkward (since they answer to different kinds of assumptions and framing devices). I indicated this by explaining that theories of distributional justice are (generally!) forward-looking whereas several of the climate ethics justice principles are (generally!) backward-looking (i.e. PPP and BPP, albeit not APP).⁵⁴ This is because distributional theories of justice give ideal shapes of justice for some metric of justice – but usually do not tell us how to get to these ideal distributions or make them dependent on historical facts about how actual distributions arose. In contrast, PPP and BPP are completely sensitive to the causal and historical sources of emissions and resources therefrom.

Again, one guiding goal was to organize rather complex debates, making them legible to audiences that were unfamiliar with these discussions. Another goal was to help translate between audiences, linking the IPCC 'equity principles' to philosophical discussions, showing that the equity principles are not very systematic and, moreover, do not come close to exhausting the theoretical normative space.

Having identified some key morally relevant issues in the carbon budgets recommendations we were developing (i.e. temperature limit, reference year, and 'equity principle'), we can now turn to how these can be evaluated depending on the level of convergence evidenced in the philosophical literature.

A.2. Evaluating Policy Options with Reference to Philosophical Literature

The first issue was about the (global) temperature limit. Here, there were legal requirements that the carbon budgeting process had to be 'consistent' with the Paris Agreement limits; that is, the 'goal of limiting global temperature increase to well below 2 degrees Celsius, while pursuing efforts to limit the increase to 1.5 degrees'. It was straightforward to point out that many climate ethicists, whether explicitly or implicitly, have supported the Paris Agreement temperature goals. The committee followed a standard interpretation of this: that a carbon budget provides at least 66% likelihood of remaining below 2°C with an

approximately 50% likelihood of remaining below 1.5°C. This is meant to reflect the plain textual meaning of committing to keeping temperature rises below 2°C and pursuing efforts to keep temperatures to 1.5°C adopted in the Paris Agreement. Obviously, these estimated probabilities do not mean that those temperature commitments will certainly be achieved, but they do represent keeping the bulk of probabilities within the committed range.

The majority of prominent philosophers tend to endorse the Paris Agreement's temperature limits, although some suggest that those limits are insufficiently stringent, especially the philosophers who focus on the attendant likelihoods.

On the critical side, Stephen Gardiner suggests that if we were discussing a student's progress (say, from the perspective of a parent), we would find those kinds of probabilities unsatisfying or even worrying.⁵⁵ Analogously, he thinks, a 66% likelihood of keeping temperatures below a threshold is not a satisfactory level of risk, writing that 'Accepting a 33–50% chance of failure hardly seems a robust commitment to protecting the future'. (Instead, he suggests that we should insist on probabilities like 95% of keeping temperatures below a target).

However, Gardiner is more critical than most moral philosophers in this discussion. Many other philosophers adopt, often with caveats, the Paris temperature limit.⁵⁶ These philosophers approvingly cite the Paris Agreement's temperature limits as having normative force, partly due to the fact that so many states have made commitments to them, although (a) often these philosophers indicate that these are minimal, not optimal, since greater mitigation would be morally superior, and (b) they rarely discuss the probabilities attached to the targets.

Some philosophers also point out that higher limits are morally objectionable.⁵⁷ For instance, Steel *et al.* point out that there is overwhelming moral reason to pursue mitigation in order to avoid pathways that could lead to 4°C by the end of the century.⁵⁸ Their reasoning is that 4°C can reasonably be construed as incompatible with stable societal functioning, and all people who can expect to live (close) to the end of the century have an overwhelming interest in avoiding societal collapse (since all reasonable conceptions of the good life rely on social stability). They conclude that it is in the interests of many currently living people to avoid pathways compatible with these temperatures, justifying mitigation efforts that lessen the likelihood of such pathways.⁵⁹

Overall, my convergent evaluation is that *Paris-aligned temperature targets are philosophically robust*, as they are consistent with a variety of professed moral philosophy positions, even if they do not necessarily command unanimity.

The second issue was the carbon budget reference year. In principle, the target relative to 2018 could be recalculated to other years, so this is partially an accounting point. But there is an important transparency issue, which is that most country reductions are relative to earlier years, so intercomparability is easier when adopting the same reference years.

The relevant philosophical consideration here is that later reference years (implicitly) grandfather in earlier distributions of emissions as well as adopt baselines that favour recent high-emitting countries, since considering the burden from a relatively high-emission economy is different from a low-emission economy.

There is almost unanimous opposition in the philosophical literature to grandfathering.⁶⁰ Summarizing the literature, Schulan *et al.* explain that it is 'typically rejected as unjust' and that, 'Generally speaking, philosophers do not defend grandfathering'.⁶¹ The reason for this is twofold: (1) with respect to wealthy and

high-emitting countries, it seems to compound the distributional injustices and reward historical lack of ambition; (2) with respect to poorer and low-emitting countries, it retards development and penalizes them for historical circumstances.⁶² These are general claims, but they tend to characterize the impacts of adopting such a principle.

Overall, if a late date, whether 2015 or 2018, practically implies substantial grandfathering, that date lies well outside the climate ethics consensus. Climate ethicists tend to use baseline years like 1990. Overall, my convergent evaluation was that the option of *dates well after 1990 as reference years is philosophically objectionable or questionable*.

The third issue is what principle of equity or distributional justice approach we adopt. In the context of (global) carbon budgets, the scope of justice is straightforward: the scope of justice is all humans or all countries (sometimes considered as single aggregated units, sometimes as sets of citizens).⁶³

The metric of justice adopted is emissions (or emissions allowances), aggregated into a global carbon budget. Note that this is not a standard metric of justice, since emissions are not in themselves morally valuable: emissions are valuable insofar as they conduce to increased consumption, welfare, or satisfaction of human needs. However, this issue can be set aside since the targets adopted in the Paris Agreement are meant to be responsive to some level of threat to one or more of these actually morally important things. That allows us to treat emissions as the metric.

This leaves only the pattern of justice: what distribution of emissions does justice allow for? Ireland adopts an egalitarian pattern, with emissions reductions relative to population. The main contention of this subsection is that, while egalitarianism is an influential⁶⁴ distributional pattern, it is only one among many – and several, if not all, of the other patterns would increase the responsibility of Ireland. For instance, if we adopted sufficientarian or utilitarian/prioritarian patterns globally or were sensitive to PPP or APP, any of these additional normative considerations would tend to increase the relative stringency for Ireland.

For these reasons, overall, my convergent evaluation found *egalitarian patterns philosophically plausible* in distributional justice terms, but egalitarian distributions are not especially sensitive to need or efficiency, so *should be thought of as conservative assumptions* (in the sense that egalitarian patterns are favourable patterns for Ireland relative to other common patterns). In other words, such a pattern is philosophically plausible, but not robust.