Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR WIRTSCHAFTS- UND
SOZIALWISSENSCHAFTEN

How to formulate climate targets under uncertainty and anticipated future learning about climate sensitivity? – An axiomatic review of the strong sustainability paradigm.

Hermann Held
Felix Schreyer

WiSo-HH Working Paper Series
Working Paper No. 54
March 2020

WiSo-HH Working Paper Series
Working Paper No. 54
March 2020

How to formulate climate targets under uncertainty and anticipated future learning about climate sensitivity? – An axiomatic review of the strong sustainability paradigm.

Hermann Held, University of Hamburg
Felix Schreyer, Potsdam Institute for Climate Impact Research

ISSN 2196-8128

Font used: „TheSans UHH“ / LucasFonts

Kontakt:
WiSo-Forschungslabor
Von-Melle-Park 5
20146 Hamburg
E-Mail: experiments@wiso.uni-hamburg.de
Web: http://www.wiso.uni-hamburg.de/forschung/forschungslabor/home/

How to formulate climate targets under uncertainty and anticipated future learning about climate sensitivity? – An axiomatic review of the strong sustainability paradigm.

Felix Schreyer[1,2], Hermann Held[2]

[1] Potsdam Institute for Climate Impact Research, Telegrafenberg, 14473 Potsdam, Germany. The major part of the work was carried out at:
[2] Research Unit Sustainability and Global Change, University of Hamburg, Grindelberg 5, 20144, Hamburg, Germany.

E-mail: felix.schreyer@pik-potsdam.de

Abstract

Strong sustanbility demands that decisions on climate mitigation be guided by a climate target and that compliance with the target be the primary concern prior to saving mitigation cost. Climate targets have often been formulated as temperature targets and for the case of uncertainty about climate sensitivity as probability targets. However, for the realistic case that we learn about climate sensitivity over the decision-making period, it is not clear how strong sustainability would consistently derive decisions on climate mitigation before and after leraning. We systematically structure the normative debate on adequate decision criteria for strong sustainability under uncertainty and learning along the lines of the von-Neumann-Morgenstern axioms of expected utility theory. We distinguish between a strict and a pragmatic-probabilistic interpretation of strong sustainability. We find that both interpretations break with the continuity axiom, while the pragmatic-proabilistic interpretation violates, in addition, the independence axiom. We discuss different possible decision criteria for strong sustainability under learning about climate sensitivity, among them a new time-recursive cost-effectiveness analysis. This probabilistic target formulation for the case of learning leads to non-trivial results if a "safe" probability level can be reached at zero mitigation cost in at least one learning scenario in which climate sensitivity turns out to be sufficiently low. This may occur if learning happens rather late and major parts of the low-carbon transformation have been achieved already before learning. Overall, our decision-analytic review helps to better understand the position of strong sustainaibility and its potential inconsistencies. We would encourage future work to use the methods of decision theory for structuring normative positions in the sustainability discourse.

Abbreviations

CBA	cost-benefit analysis
CEA	cost-effectiveness analysis
CRA	cost-risk analysis
PP-CEA	Posterior-Prior cost-effectiveness analysis
EU	expected utility

Introduction

The notion of a climate target is at the heart of global climate policy. The Rio Conference in 1992 agreed on stabilizing "greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system" (UNFCCC 1992). Eighteen years later at the climate conference of Cancún, this level was specified as a warming of 2°C global mean temperature relative to preindustrial times and reaffirmed in the legally binding Paris agreement of 2015 signed by 196 countries (UNFCCC 2011; UNFCCC 2015).

A global climate target has often been understood as an implementation of strong sustainability, a school of thought who maintains that certain forms of the natural capital at stake cannot be substituted by human-made capital (Neumayer 2013). The target level is considered a maximum acceptable limit whose transgression cannot be compensated by gains in other areas (WBGU 2011; WBGU 2014). Contrary to weak sustainability, externalities from climate change are not absorbed into a welfare functional but imposed as constraints to welfare maximization.

Much work has been done on analyzing economic transformation pathways to efficiently reach various climate stabilization levels (IPCC 2014, chap. 6). Without uncertainty, a welfare functional is typically maximized subject to a constraint on global greenhouse gas concentrations or global mean temperature. This is known as a cost-effectiveness analysis (CEA) of the climate target. Taking uncertainty about the climate response to emissions into account, a probabilistic climate target can be formulated, for instance, as keeping global temperature below 2°C with a probability of at least 66%. CEA of probabilistic climate targets has been conducted first by den Elzen & Van Vuuren (2007) and Held et al. (2009) and is the common approach to scenario analysis of climate stabilization today (IPCC 2018) .

But, how can a climate target be formulated under uncertainty and learning, i.e. if future mitigation decisions can be adapted in the light of new information about the climate response to emissions? Webster et al. (2008) estimate that the uncertainty about climate sensitivity will be reduced by 20-40% over the next one to four decades by Bayesian learning from climate observations. Moreover, advances in the conceptual understanding as, for example, in cloud physics may reduce this uncertainty (IPCC 2013, pp. 593-594). Learning implies that a transformation pathway has a different 2°C exceedance probability depending on the state of knowledge (probability distribution) about climate sensitivity. So far, there has been no formulation of a (probabilistic) climate target for this case. Schmidt et al. (2009; 2011) discard different forms of CEA with learning due to consistency problems with probabilistic constraints.

Instead, they propose cost-risk analysis (CRA) which, as an expected utility (EU) criterion, satisfies common consistency principles. CRA has been used to investigate optimal mitigation pathways for the case of future learning about climate uncertainty (Neubersch et al. 2014), delayed climate policy (Roth et al. 2015) and climate engineering (Roshan et al. 2018). Yet, as an EU criterion, this approach is an unconstrained welfare maximization and thus at odds with the strong sustainability paradigm.

The aim of this paper is twofold: Our main question is how a climate target can be formulated for the case of climate-related uncertainty and learning. We tackle this question by a systematic review of axioms in decision making against the background of the climate problem, linking the discourse on sustainability and climate targets to the foundations of decision theory. We are not aware of such methodological links in the literature on climate mitigation. A second objective is thus to explore this interdisciplinary perspective and demonstrate the usefulness of axiomatic methodology for structuring the sustainability debate.

The analysis is structured as follows: Based on a literature review, the first section introduces strong sustainability as the normative reasoning behind climate targets. Second, we present cost-risk analysis, the criterion used to derive target-based decisions under learning so far, and its conflict with strong sustainability. Third, to open and structure the space of possible decision criteria, we discuss the necessity of complying with each of the von-Neumann-Morgenstern axioms of EU theory in the context of the climate problem. We identify different classes of eligible decision criteria conditional on set of axioms a proponent of strong sustainability is

willing to accept. Fourth, based on two interpretations of strong sustainability presented in the first section, we review different formulations of a CEA under uncertainty and learning proposed by Schmidt et al. (2009; 2011) and, moreover, suggest a new time-recursive CEA. Finally, we summarize the proposed criteria and discuss advantages and limitations of lexicographic criteria relative to EU criteria for making decisions on the climate problem under uncertainty and learning.

1. Strong Sustainability: The Reasoning behind Climate Targets

In the sustainability discourse, there two competing paradigms of how to approach an environmental problem: weak and strong sustainability. Neumayer (2013) provides a comprehensive review of the broad debate on the two concepts. Fundamentally, they differ on whether the natural capital at stake in the environmental problem (e.g. a forest area, natural resources, the ozone layer or the state of the global climate) is substitutable by human-made capital.

According to Neumayer (2013), weak sustainability requires that total net investment be positive or at least zero, i.e. the aggregate stock of capital, human-made capital and natural capital, should be non-declining. Strong sustainability makes the additional requirement on the stock of natural capital. There are two versions: Either the aggregate natural capital should be maintained in value terms or certain stocks of "critical" natural capital should be maintained in physical terms. The key difference between weak and strong sustainability is the substitutability assumption of natural capital which generates two fundamentally different perspectives. While the weak sustainability is concerned with the adequate pricing of natural capital relative to human-made assets, strong sustainability seeks to impose maximum acceptable limits of environmental stress that should not be transgressed.

The body of literature on both paradigms and their specifications is enormous. The foundations of weak sustainability were laid by Robert Solow and John Hartwick (Solow 1974; Hartwick 1977). Their underlying substitutability hypothesis features the standard approach of cost-benefit analysis presented in environmental economics textbooks (e.g. Perman et al. 2003, pp. 351). The weak sustainability paradigm has been criticized early and fiercely, for instance, by Georgescu-

Roegen (1975) and Daly (1974; 2007) who maintained that there are physical limits to the size of a sustainable economy. Whether as "safe minimum standards" (Ciriacy-Wantrup 1952), "optimal scale" (Daly 1992, 2005), "tolerable windows" (Petschel-Held et al. 1999), "planetary boundaries" (Rockström et al. 2009) or "planetary guard rails" (WBGU 2011), the idea of environmental limits has been very influential especially with respect to global problems. Over the last three decades, numerous authors have contributed to the broad debate on weak and strong approaches which only step by step let go of universal claims for the insight that, depending on the environmental problem, the substitutability assumption may be context-specific (Neumayer 2013).

In climate change economics, the very controversy appears between proponents of cost-benefit analysis (CBA) and proponents of a cost-effectiveness analysis (CEA) of climate targets. CBA weighs climate damages in monetary terms against mitigation cost. Yet, estimations of climate damages as provided by Nordhaus (2008; 2013) or Tol (2002; 2009) have been strongly criticized mainly on two grounds. First, the manifold impacts of climate change on human well-being are fundamentally uncertain and hard to quantify and, second, their monetary valuation must rely on ethically contestable methods and assumptions (Ackerman et al. 2009; Charlesworth & Okereke 2010; Pindyck 2013). Instead, proponents of climate targets have argued along the lines of the precautionary principle: As long as our knowledge is as limited, they claim, it is best to stay in relatively familiar and safe climatic range. The case for approaching the climate problem by maximum acceptable limits instead of internalizing climate damages into an economic welfare optimization has been made repeatedly (e.g. WBGU 1995; Schellnhuber 1998; Ackerman et al. 2009; Rockström et al. 2009; Neumayer 2013). Although reference to the two terms and their long-standing literature have become sparse, weak and strong sustainability are very present in the climate change debate.

The priority of holding the critical limit is key to strong sustainability. The WBGU (2011, p. 32) considers the 2°C limit of global warming as one of the

> *"damage thresholds whose transgression either today or in the future would have such intolerable consequences that even large-scale benefits in other areas could not compensate these."*

Introducing the 2°C target to climate policy, the WBGU (1995) argued that a warming within 2°C relative to preindustrial would leave the planet in a climate state relatively familiar from paleo-climatic evidence of the past 800,000 years. Beyond that, we would enter a climatic range never experienced by human beings with potentially disastrous large-scale changes on our planet. Other authors have argued similarly for a strict limit on the basis of precaution although with different emphasis on how much is known about the impacts in case of transgression (see Neumayer 2013, pp. 40-46).

This value system corresponds to lexicographic preferences that follow an order of decision criteria. First, a primary criterion ("not transgress the guard rail") is applied. If the primary criterion is not decisive, a secondary criterion ("large-scale benefits in other areas") is applied, and so on. We will consider strong sustainability as demanding lexicographic preferences for, first, reaching the climate target and, second, minimizing economic mitigation cost. This corresponds to a cost-effectiveness analysis (CEA) of the climate target.

As a lexicographic criterion, strong sustainability demands attaining the environmental target at any cost. However, what if the cost become very large? As Neumayer (2013, pp. 124-126) points out, two interpretations of strong sustainability have evolved on this matter: First, ignoring opportunity cost is a deliberate decision since transgressing the environmental limit is, in fact, the worst that can happen. Second, costs are considered implicitly when the environmental target is developed. The precondition is that the target must not incur unacceptably high cost. Here, strong sustainability can be understood as recommending a precautionary low-cost-low-risk option in the face of fundamental uncertainty, although this option might with more knowledge turn out not to be the optimal choice. The WBGU (1995) added that the 2°C target was only reasonable because it would not impose "excessive cost" to the global economy. However, other publications do not consider mitigation cost for setting a climate target (WBGU 2011; Rockström et al. 2009). We will take into account both interpretations by distinguishing between a strict target and a pragmatic-probabilistic target where the latter allows for some limited exceedance probability to avoid excessive mitigation cost.

One might argue that the uncertainty about climate sensitivity only adds to the uncertainty about climate impacts by which the temperature limit was justified in the first place. We should therefore formulate the target not in terms of temperature but of a variable over which we have

sufficient control. However, this ignores that the uncertainty about climate impacts is more fundamental than, for instance, uncertainty about climate sensitivity which can be reasonably quantified by probabilities (IPCC 2013, pp. 921). The temperature limit separates those two domains of uncertainty and allows target-based decision analysis to work probabilistically. This allows us to model the more realistic case that one type of uncertainty, the uncertainty about climate sensitivity, will be reduced over time (learning) by future observations (Webster et al. 2008), while the other type, the impact uncertainty, prevails longer.

2. Cost-Risk Analysis: A Target-based Expected Utility Criterion

Cost-risk analysis (CRA) has been suggested and applied as a possible target-based decision criterion for learning (Schmidt et al. 2011; Neubersch et al. 2014). To formalize this and the following decision criteria we exemplarily consider uncertainty about climate sensitivity θ known up to a prior probability density distribution $p(\theta)$. In the static case, i.e. without learning, we consider cumulative global greenhouse gas emissions E as decision variable resulting in a maximum global temperature over time $T(E, \theta)$ measured relative to preindustrial temperature. This is a common simplification of the problem since maximum temperature is approximately proportional to cumulative emissions such that the timing of emissions is less important (Allen et al. 2009). Moreover, we consider aggregate economic mitigation cost $C(E)$ incurred relative to a business-as-usual growth scenario without climate damages.

With learning, the decision problem becomes dynamic, i.e. there are two stages in the decision process: a first-period decision before learning and a second-period decision after learning. We consider n possible learning scenarios (messages) with posterior distributions $\boldsymbol{p}(\theta) = (p_1(\theta), \dots, p_n(\theta))$. The learning scenarios are obtained with prior probabilities $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$, where $\sum_m \pi_m = 1$. Together, they form an information structure $(\boldsymbol{\pi}, \boldsymbol{p}(\theta))$. Bold notation denotes vectors over learning scenarios. The decision maker decides on the first-period emissions E_0 and the n second-period emissions $\boldsymbol{E_m} = (E_1, \dots, E_n)$ of all learning scenarios. Our decision variable is therefore the tuple $(E_0, \boldsymbol{E_m})$.

Cost-risk analysis (CRA) finds the optimal $(E_0, \boldsymbol{E_m})$ by minimizing a weighted sum of mitigation cost and climate risk:

$$Min_{((E_0,\boldsymbol{E_m}))}\ \varepsilon_m \varepsilon_{\theta|p_m(\theta)}\ [C(E_0, E_m) + \beta\ X(T(E_0, E_m, \theta), T^*)]. \quad (1)$$

Here, X is an exceedance measure of the temperature target T^* and β is a trade-off parameter that represents the willingness-to-pay for preventing a unit exceedance. The operators $\varepsilon_m[.]$ and $\varepsilon_{\theta|p_m(\theta)}[.]$ denote the expectation over the learning scenarios and over climate sensitivity given the posterior $p_m(\theta)$. Climate risk is some functional of the distribution of exceedance. The difference to CBA is that X is not a climate damage function based on aggregating specific impact evaluations but a function increasing with temperature overshoot that represents the decision maker's aversion to the exceedance of the critical limit.

Although a target-based criterion, cost-risk analysis clearly conflicts with the above rationale of strong sustainability since it is not a lexicographic criterion and allows for compensating exceedance as soon as sufficient mitigation cost can be saved. However, we will see in the following section that, as an EU criterion, it aligns with a number of common consistency principles.

3. Explaining and Discussing the von-Neumann-Morgenstern Axioms

Since the foundational work by von Neumann and Morgenstern (1944) expected utility (EU) theory has become the standard framework for decision-making under uncertainty. They show that, as soon as a decision maker accepts four general consistency principles, the von-Neumann-Morgenstern axioms, she finds her optimal choice by maximizing an expected utility function. We will explain and discuss each of axioms in the context of the climate problem. This serves to structure the discussion on possible decision criteria of strong sustainability under uncertainty and learning.

The von-Neumann-Morgenstern framework conceives decision under risk as a choice between lotteries. Risk implies that the space of possible states of the world and their probabilities are known. As depicted in Figure 1, there are three types of lotteries: simple lotteries, compound lotteries and dynamic lotteries (e.g. Machina 1989). Simple lotteries are one-stage bets and

correspond to probability distributions over the possible outcomes (Figure 1a). In case of the climate problem without learning, by choosing emissions E under the probability distribution $p(\theta)$ we obtain a simple lottery $L_{E|p(\theta)}$ on outcomes which are temperature-cost pairs $\{T(E,\theta), C(E)\}$. Compound lotteries are two-stage lotteries on simple lotteries (Figure 1b). The compound lottery $p\ L_1 + (1-p)L_2$ is the lottery to receive the simple lottery L_1 with probability p and the simple lottery L_2 with probability $(1-p)$. We moreover assume the axiom of reduction which implies that any compound lottery can be reduced to a simple lottery by multiplying the probabilities along the paths in the decision tree. Finally, a dynamic lottery is a two-stage lottery on decisions between simple lotteries (Figure 1c). In the case of learning, the first-period decision E_0 is a choice between dynamic lotteries on different second-period decision problems. Each of the second-period decisions $\boldsymbol{E_m}$ is a choice between simple lotteries.

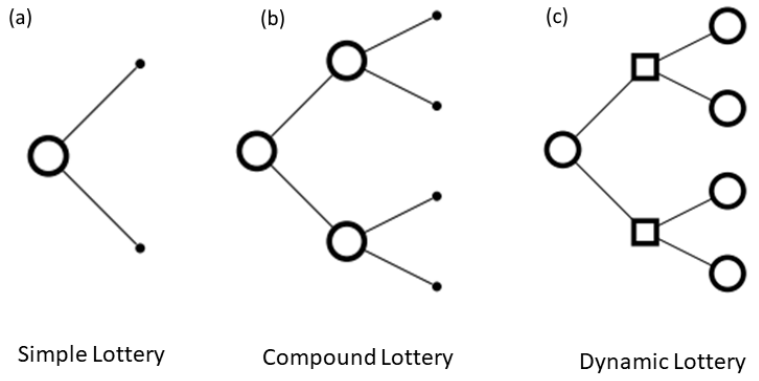

Figure 1: Decision trees of simple, compound and a dynamic lotteries. Circles denote lottery nodes (chance decides), squares denote decision nodes (decision maker decides) and black points denote outcomes (payoffs).

Decision analysis investigates the consistency of a set of preferences a decision maker holds when asked about pairwise comparisons of lotteries. The strict relation $L_1 \succ L_2$ denotes that she prefers the lottery L_1 over the lottery L_2 and $L_1 \sim L_2$ implies that she is indifferent between the two. The weak relation $L_1 \succcurlyeq L_2$ denotes that she holds either a preference $L_1 \succ L_2$ or an indifference $L_1 \sim L_2$.

The lottery space of simple lotteries Λ that we consider includes all combinations of emissions $((E_0, E_m)$ with one of the posterior distributions in $\boldsymbol{p}(\theta)$ as well as the reductions of all possible compound lotteries with the likelihoods $\boldsymbol{\pi}$ on those first lotteries. The former correspond to the second-period option space (after learning). As we will see later in this section, the latter

correspond to the first-period option space (before learning) provided that two principles of dynamic choice are accepted.

Moreover, we assume that the posterior distributions in $\boldsymbol{p}(\theta)$ have infinite support, i.e. although its probability may be low, an arbitrarily high value of climate sensitivity cannot be ruled out. This is consistent with the distributions given by the IPCC (2013, pp. 1107). They estimate a 90% probability for (equilibrium) climate sensitivity to be below 6°C. However, the complexity of the climate system with its numerous feedback mechanisms does not allow for constraining climate sensitivity to a maximum level. Learning about climate sensitivity in the next decades will not find an upper bound either as long as observations come with infinite support.

After defining the relevant lottery space Λ for the climate problem under uncertainty and learning, we present the four von-Neumann-Morgenstern axioms following Gollier (2001, pp. 4-6):

(I) **Completeness:** *Preferences* $\succcurlyeq$ *on the lottery space* Λ *are such that for any two lotteries* $L_1, L_2 \in \Lambda$ *it is either* $L_1 \succ L_2$, $L_1 \prec L_2$ *or* $L_1 \sim L_2$.

Completeness demands from the decision maker to compare all available lotteries pairwise and state a preference. She must either prefer one option over the other or be indifferent between the two. There is no third category. While the other axioms deal with consistency between lottery preferences, this one ensures that there are well-defined preferences in the first place.

Completeness over the space of simple lotteries Λ is certainly demanding for a decision problem as complex as the climate problem. However, the axiom is necessary for the existence of an optimal choice on all subsets of the lottery space $S \subseteq \Lambda$**.** An optimum on Λ might also exist for incomplete preferences, yet adding empirical constraints to the problem can lead to infeasibility on a smaller subset. Past emissions, for instance, prescribe a minimum temperature increase regardless of the mitigation decision we make.

In a normative context of assessing different future scenarios of climate change mitigation, we want a reasoned choice (Gilboa 2009, pp. 131-132). Preferences cannot be observed, they need to be justified. Yet, already by asking about the preferences between two outcomes, i.e. temperature-cost pairs, the decision maker might find it difficult to develop reasoned preferences

in the light of high temperatures or "excessive" mitigation cost. The corresponding predictive uncertainties and moral trade-offs may be overwhelming such that a pair of outcomes could be considered "incomparable".

Now, for a proponent of the strict target, a transgression of the critical temperature is inacceptable regardless of the mitigation cost, so she has no difficulties in stating complete preferences. A proponent of the pragmatic-probabilistic target who makes cost considerations prior to defining the target level, though, may seek to avoid "tragic choices"[1] between high mitigation cost and high climate risk because of the additional effort required in her decision-making process. From a perspective of bounded rationality, we argue that climate targets are preferable to cost-benefit analyses. Developing preferences over tragic choices requires not only a better understanding of a world with high degrees of warming or large mitigation challenges. Moreover, unequally distributed global mitigation cost and climate risks will make it even more difficult to negotiate tragic choices in international agreements on climate policy. A decision maker averse to making tragic choices will aim to circumvent the completeness axiom on the whole space of possible temperature-cost combinations.

The gist of strong sustainability is that if the climate target level can be reached at low cost, a "satisficing" solution is already found and preferences between options beyond the target level are not necessarily needed. In the light of fundamental uncertainty, the approach does not look for an overall optimal pathway, but for a safe pathway which is the essence of the precautionary principle at the basis of strong sustainability (IPCC 2014, pp. 172). As long as cost-effectiveness analyses of a strict or a probabilistic climate target without learning are feasible, these decision criteria have the advantage that they do not require complete preferences.

However, considering the climate problem with learning, "tragic choices" cannot be ruled out. As past emissions have already occurred and the prior distribution of climate sensitivity has infinite support, there is always a small chance of ending up in a very "bad" learning scenario with the choice between high mitigation cost and high climate risk. Any temperature or exceedance probability of that temperature may be transgressed after learning if only we look at a case of sufficiently high climate sensitivity. Since, in the case of learning, the decision maker

[1] A term used by Edenhofer and Lessman (2007).

needs to anticipate her actions after learning, proponents of strong sustainability must also structure the preference space beyond the target level. Completeness on the whole set of mitigation cost and climate risk combinations becomes necessary to ensure the existence of an optimal choice.

(II) **Transitivity:** *Preferences* $\succcurlyeq$ *on the lottery space* Λ *are such that for any* $L_1, L_2, L_3 \in \Lambda$*:* $L_1 \succcurlyeq L_2$ *and* $L_2 \succcurlyeq L_3$ *implies* $L_1 \succcurlyeq L_3$.

Transitivity is consistency over a triple. It is readily compelling as soon as the decision maker can clearly tell the lotteries apart[2]. Complete and intransitive decision makers can be exploited by "money pumps" (e.g. Mandler 2005): Assume the decision maker prefers $L_1 \succcurlyeq L_2$ and $L_2 \succ L_3$, then she would be willing to take some disadvantage (e.g. paying money) to trade L_2 for L_3. Also, she would not mind exchanging L_1 for L_2. Now, if she preferred $L_3 \succ L_1$ she would trade L_1 for money to obtain L_3, which is the lottery she held in the beginning. This procedure can be used to "pump" an infinite amount of money (disadvantage) out of a complete and intransitive decision maker which we see as inacceptable from a social planner perspective.

For a finite number of options in the lottery space, completeness and transitivity allow to pairwise compare all lotteries and arrange them on a scale from "worst" to "best". This constitutes an ordinal utility function, that is, a utility ranking where utility differences do not have a meaning except for that an option with higher utility is preferred over an option with lower utility. An ordinal utility difference does not give information on how "much more" preferred an option is over another (Gilboa 2009, pp. 53-54).

To construct a utility function for the problem with infinitely many options, we need to accept a third axiom:

(III) **Continuity:** *Preferences* $\succcurlyeq$ *on the lottery space* Λ *are such that for any* $L_1, L_2, L_3 \in \Lambda$ *with* $L_3 \succcurlyeq L_2 \succcurlyeq L_1$ *there exists a probability* $p \in [0,1]$ *such that:*
$pL_1 + (1-p)L_3 \sim L_2$.

[2] For measurement problems if the decision maker cannot tell two different options apart and the corresponding theory of semiorders, see Gilboa (2009, pp. 65-71).

Let us elucidate the definition with an example: Imagine an adventurer on a treasure hunt. She arrives at a bridge over a canyon and has to decide whether or not to dare the crossing. There are three possible outcomes: First, she manages to cross the bridge and finds the treasure (T). Second, she falls off the bridge and dies (D) or, third, she goes back without a treasure (B). Of course, she will prefer $T \succ B \succ D$. The decision of whether or not to cross is between the two lotteries $pT + (1-p)D$ and B. Here, $p \in [0,1]$ is the probability that she safely crosses the bridge, the stability of the bridge. If p is high (a concrete bridge), she would cross, while for a low probability (a rope in midair), she would prefer going back. Now, continuity demands that there exist some kind of bridge for which her decision of whether or not to cross will be quite hard, i.e. an break-even probability $p \in [0,1]$ at which she is indifferent: $pT + (1-p)D \sim$ B.

The continuity axiom can be understood in analogy to mathematical continuity (Figure 2): A small change in the probabilities underlying two lotteries should only make a small change in the preferences over them. Abrupt changes, where for an arbitrarily small increment in probability the decision maker swaps from a strict preference to the opposite strict preference, are not allowed. Instead, there must be a smooth transition over an indifference relation.

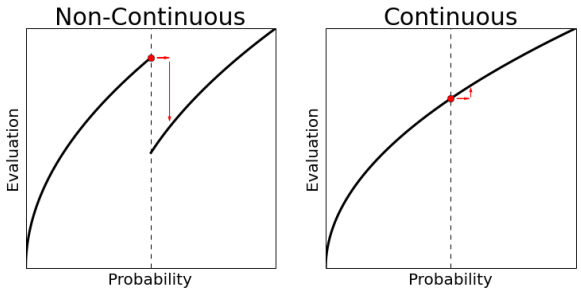

Figure 2: Illustration of non-continuity and continuity. Continuity demands that a small change in the probabilities implies only a small change in the evaluation of a lottery.

Again, we distinguish between the strict and the pragmatic-probabilistic interpretation. The discontinuity induced by the strict target, i.e. the probability threshold at 100%, captures the idea of the precautionary principle that the certainty of preventing “intolerable damage” has a different quality than a mere arbitrarily high probability. Especially since the climate problem is global and intergenerational such that those who induce the risk do not necessarily bear the risk, this distinction between no risk and a small risk of “intolerable damage” is valid. However, if climate sensitivity cannot be constrained to an upper bound and past emissions have already occurred, it is impossible to stay below the temperature limit with certainty.

A probabilistic temperature target that specifies a maximum acceptable exceedance probability larger than zero may be feasible and economically affordable, yet its justification for breaking with the continuity axiom is not clear. The 17th Conference of the Parties in Durban established the notion of holding a "likely chance" to reach the 2°C target (UNFCCC 2012) which has been interpreted as a probability of 66% (Neubersch et al. 2014). Unlike the 2°C target, such probabilistic target is not based on historic or predictive insight of climate science. It is the result of a policy process, not an assessment of critical environmental limits. We find no reason why a small probability increase at a specific non-zero exceedance probability should be disproportionately more dangerous than an increase at any other probability level.

Accepting completeness, transitivity and continuity on the lottery space Λ implies the existence of a real-valued ordinal utility function (Gollier 2001, pp. 5-6): Completeness and transitivity let us find a best lottery $\overline{L}$ and a worst lottery $\underline{L}$ and by continuity we find for any $L \in \Lambda$ a unique probability $p \in [0,1]$ such that $L \sim p\overline{L} + (1-p)\underline{L}$. The probability p can then be interpreted as an ordinal utility representation of the lottery L. Thus, for the static problem (simple lotteries) by complying with the first three axioms, there exists a continuous utility function $V\big(E\,, p(\theta)\big)$ that ranks our options of emissions E for a given probability distribution of climate sensitivity $p(\theta)$ from best to worst. For finding the optimal choice, we perform a utility maximization

$$Max_{(E)}\ V\big(E\,, p(\theta)\big). \tag{2}$$

This is the general form of a decision criterion a proponent of strong sustainability would reject. Since temperature and mitigation cost are both functions of E some trade-off function needs to be defined that relates the two to each other. This allows, in principle, for any climate target exceedance if mitigation cost are sufficiently high. A proponent of strong sustainability must therefore drop either completeness, transitivity or continuity.

The fourth axiom is

(IV) Independence: *Preferences* $\succcurlyeq$ *on the lottery space* Λ *are such that for any* $L_1, L_2, L_3 \in \Lambda$ *and* $p \in [0,1]$: $L_1 \succcurlyeq L_2 \Leftrightarrow pL_1 + (1-p)L_3 \succcurlyeq pL_2 + (1-p)L_3$.

Independence demands that the preferences on $pL_1 + (1-p)L_3$ and $pL_2 + (1-p)L_3$ be determined by the preferences over L_1 and L_2 regardless of what p and L_3 are. Wakker (1999)

decomposes the independence axiom into three principles of dynamic choice, i.e. choice that involves dynamic lotteries: consequentialism, time-consistency and context-independence[3]. They are illustrated in Figure 3 where each consistency principle implies that the decision maker takes the same decision in two neighboring decision trees, i.e. goes for the upper/lower branch in both problems.

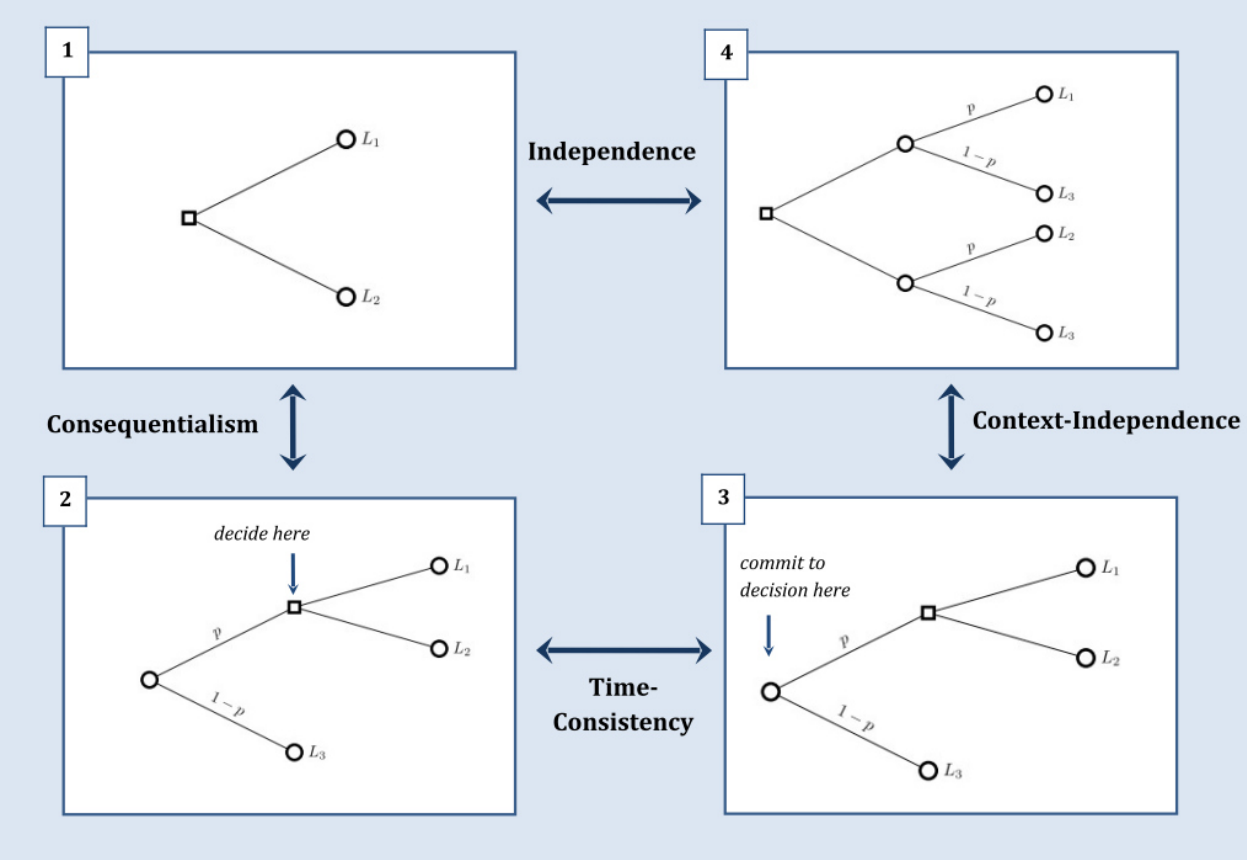

Figure 3: Relation between independence, consequentialism, time-consistency and context-independence following Wakker (1999) *and Gollier* (2001, p. 12). *Each consistency principle implies that the decision maker takes the same decision at the decision node (square) in two neighboring decision trees, i.e. goes for the upper/lower branch in both problems.*

Consequentialism ("foregone-event independence" for Wakker) implies that the decision should not depend on what could have happened in the past but eventually did not occur. Time-consistency requires that the decision maker can correctly anticipate her future choice. Context-independence holds that a time-consistent decision maker can frame a dynamic lottery as a compound lottery by committing to the anticipated choice before learning. Finally, choosing in

[3] Wakker also includes axiom of reduction which we take for granted at this stage.

panel (1) as in panel (4) corresponds to the independence axiom. To reject independence, either consequentialism, time-consistency or context-independence must be dropped.

Sparked by the famous Allais Paradox (Allais 1953), a considerable body of literature exists on violations of the independence axiom and corresponding Non-EU theories in descriptive as well as normative contexts (e.g. Kahneman & Tversky 1979; Loomes & Sugden 1982; Machina 1989; Bradley & Stefansson 2016). Wakker (1999) categorizes the arguments made against independence into arguments against consequentialism, time-consistency or context-independence. We will briefly discuss each of them in the context of the climate problem.

First, breaking with consequentialism would imply that the decision after learning takes into account counterfactual learning scenarios that have not realized. It would matter for the decision after learning whether or not we end up with a "good" or "bad" learning scenario relative to the other learning scenarios that were possible. As a collective and intergenerational issue, though, the climate problem should be tackled forward-looking and consequentialist, we think. It is dangerous giving future decisions makers the possibility to justify their mitigation decisions by pointing to decisions they would have made in counterfactual learning scenarios which makes the discourse somewhat irrational because such statements cannot be disproven.

Second, time-inconsistency would allow the decision maker's plans before learning to deviate from the actual decision made after learning. In a non-strategic social planer context, this does not make sense. Correct anticipation of future choice is indispensable since the purpose of learning in time is to adjust choices conditional of different possible learning scenarios.

Third, context-independence reduces the dynamic decision problem to a static one since it maintains that a time-consistent decision maker should consider a dynamic lottery (Figure 3.3, bottom right) as a compound lottery (Figure 3.4, top right). It implies that, as the decision maker knows all possible posterior distributions and thus also her respective second-period option spaces, it does not matter whether she commits to the second-period choice today (compound lottery) or only after learning (dynamic lottery). Decision and lottery nodes are interchangeable in order.

The principle ensures that the decision maker does not reject costless learning (Wakker 1988): By choosing the same second-period emissions in all learning scenarios $\boldsymbol{E_m} = (E_1, \dots, E_1)$, she

will not be better or worse-off than as if she had chosen E_1 in the static problem without learning. The argument is that, if she does not adjust her second-period decisions, her prior probability distribution (lottery) over the temperature-cost outcome space is the same as if she had chosen the sum of the first-period and second-period cumulative emissions in the no-learning case. Context-independence disregards the specific posterior exceedance probabilities as long as they sum up to the same prior.

However, a violation of context-independence might be in the sense of strong sustainability. Meeting a probabilistic climate target (e.g. a 66% chance to stay below 2°C) under several (also more pessimist) probability distributions may be preferred to meeting it under only one probability distribution of climate sensitivity. Although the decision maker might be worse-off with costless learning in terms of mitigation cost as Schmidt et al. (2009, 2011) point out, she is better-off in terms of climate risk as the probabilistic target becomes more ambitious if referred to more than one probability distribution. Since two metrics, climate risk and mitigation cost, need to be compared, the argument that a proponent of strong sustainability would reject costless learning is not valid. Moreover, non-independence allows the decision maker's ambition level (the willingness to pay mitigation cost for reducing climate risk) to increase the more probable the exceedance of the critical temperature becomes because utility functions can be non-linear in probabilities (Gollier 2001, pp. 10-12). This property is interesting for a proponent of strong sustainability since it allows her to spend maximum mitigation cost in worst-case learning scenarios, yet without having to go for maximum mitigation already before learning in the anticipation of these scenarios. We will come back to this property in the next section when discussing Lexicographic EU criteria.

Let us summarize the result of the von-Neumann-Morgenstern framework: Accepting the independence axiom in addition to the first three axioms imposes the utility function $V\big(E\,, p(\theta)\big)$ to have a certain form: it must be linear in the probability distribution $p(\theta)$ (Gollier 2001, pp. 10-12). The axioms (I) to (IV) are the necessary and sufficient conditions for EU maximization over the space of simple lotteries which reads:

$$Max_{(E)}\ \varepsilon_{\theta|\,p}[U(E,\theta)]. \tag{3}$$

Here, $U(.)$ is a utility function of the temperature-cost outcomes $\{T(E,\theta), C(E)\}$ obtained from emissions E and climate sensitivity θ.

Since EU criteria are time-consistent and context-independent, dynamic lotteries can be reduced to simple lotteries. This allows applying EU criteria also to the two-stage decision problem in the case of learning:

$$Max_{(E_0, \boldsymbol{E_m})}\ \varepsilon_m \varepsilon_{\theta|\, p_m}[U(E_0, E_m\ , \theta)]. \tag{4}$$

To structure the discussion on alternatives to EU maximization, Table 1 presents classes of decision criteria that are compatible with different configurations of continuity and independence given that the first two axioms hold. As shown above, continuity implies a utility maximization either with linear probabilities (expected utility) or non-linear probabilities (non-expected utility). Blume et al. (1991) show that a lexicographic expected utility criterion, where multiple expected utility functions are maximized lexicographically, satisfies independence but breaks with the continuity axiom[4]. The notation $Lex\ Max\ \{V_1(.), V_2(.), \dots\}$ implies that we maximize $V_1(.)$ first, and for equal levels of $V_1(.)$, we maximize $V_2(.)$ and so on, until we obtain a complete ordering. The most general framework, violating both continuity and independence, is

	Continuity	**Non-Continuity**
Independence	$Max\ \varepsilon_{\theta\|p}[U(.\,|\theta)]$ **expected utility** (necessary and sufficient)	$Lex.Max\ \{\varepsilon_{\theta\|p}[U_1(.\,|\theta)], \varepsilon_{\theta\|p}[U_2(.\,|\theta)], \dots\}$ **lexicographic expected utility** (necessary)
Non-Independence	$Max\ V(.\,|p(\theta))$ **non-expected utility** (sufficient)	$Lex.Max\ \{V_1(.\,|p(\theta)),\ V_2(.\,|p(\theta)), \dots\}$ **lexicographic non-expected utility** (necessary)

Table 1: Summary of the classes of decision criteria compatible with different positions on the continuity and independence axiom given that completeness and transitivity are satisfied. The brackets state whether this combination of axioms is necessary, sufficient or both for the corresponding class of decision criteria.

[4] Note that (I), (II) and (IV) are only necessary but not sufficient for lexicographic EU maximization. The additional assumptions are minor though, see Blume et al. (1991).

lexicographic non-expected utility maximization where we allow for utility functions in the lexicographic structure that are non-linear in the probabilities.

4. Strong Sustainability under Uncertainty and Learning

Strong sustainability implies using a lexicographic criterion where the primary criterion is to meet a climate target. Let us now go through specific criteria that we could offer such proponent. First, the strict interpretation of a climate target can use a Lexicographic EU criterion as it complies with axioms (I), (III) and (IV) but violates continuity. We suggest

$$Lex.\ Min_{(E_0,\boldsymbol{E_m})} \begin{Bmatrix} \varepsilon_m[P(E_0,E_m,p_m)], \\ \varepsilon_m[C(E_0,E_m)] \end{Bmatrix}, \tag{5}$$

where $P(E_0,E_m,p_m)$ is the posterior probability to exceed the critical temperature T^*. It depends on first-period emissions E_0, the second-period emissions E_m and the posterior distribution of climate sensitivity $p_m(\theta)$. The criterion minimizes the upper function that represents the prior exceedance probability first, and the lower function, expected mitigation cost, second. As for all criteria that follow, the static case without learning can be obtained by setting all posterior distributions to the prior.

The obvious problem of the strict target approach is that it always suggests maximum emission reduction if the prior climate sensitivity distribution is unbounded. Any primary lexicographic function that is linear in $p(\theta)$ and strictly increasing with excess temperature above the target level (as for example the different risk measures used by Neubersch et al.) will result in such corner solution that simply ignores mitigation cost.

An interesting result of our analysis is that the intuition behind pragmatic-probabilistic strong sustainability may, in fact, be more intimately linked to the violation of independence than to the violation of continuity. Independence requires the willingness to pay for reducing exceedance probability (or climate risk in general) by one unit to be constant with exceedance probability, while continuity requires it to be continuous. The statement, though, that we should not care too much about low exceedance probabilities, but invest everything once exceedance probabilities become close to 1 is, above all, a break with the independence axiom.

The pragmatic-probabilistic interpretation minimizes mitigation cost as long as the exceedance probability is below some acceptable limit P^*. This is what is known under Chance Constrained Programming (Held et al. 2009) or Probabilistic CEA (Schmidt et al. 2009). As learning requires complete preferences on the lottery space Λ if the prior $p(\theta)$ is unbounded, we define preferences beyond the target level following the general framework of *strong sustainability under uncertainty* by Baumgärtner and Quaas (2009). That is, if a transgression of the probability limit cannot be avoided, we demand minimizing the overshoot and keep the exceedance probability as low as is still possible. The risks beyond the critical level are not traded against lower mitigation cost which preserves the key idea of strong sustainability.

The question arises to what state of knowledge (prior or posterior) the probability limit applies when making first-period and second-period decisions. We discuss three suggestions: Posterior-CEA, Prior-CEA and Posterior-Prior-CEA all of which violate the continuity and independence axiom.

Posterior-CEA:

Posterior-CEA was already discussed by Schmidt et al. (2009; 2011) and we write it in complete form as

$$Lex.\ Min_{(E_0,\boldsymbol{E_m})} \begin{Bmatrix} \varepsilon_m[\Theta[P(E_0,E_m,p_m)-P^*]P(E_0,E_m,p_m)], \\ \varepsilon_m[C(E_0,E_m)] \end{Bmatrix}. \quad (6)$$

This target formulation places the threshold P^* on the posterior exceedance probability. It chooses cost-efficient first-period and second-period emissions such that the transgression of the probability threshold P^* is as small as possible under posterior knowledge in each of the learning scenarios. Posterior-CEA violates context-independence since by choosing the same second-period emissions in all learning scenarios we do not necessarily recover the preference order over the emissions implied by the criterion without learning, i.e. if all posteriors correspond to the prior. As pointed out above, this is consistent with a position that favors holding the probabilistic target also under possible future posteriors. However, the problem with Posterior CEA is its extreme anticipation effect: The decision maker would pay any amount of mitigation cost in the first-period only to further reduce the exceedance probability in one worst-case learning scenario

which may even not be very likely. She ignores the likelihoods of these scenarios and is completely fixated on the possible worst-case which, in addition, makes the first-period decision very sensitive to the sample of learning scenarios considered.

Prior-CEA:

Schmidt et al. (2009; 2011) propose another criterion that reduces to Probabilistic CEA without learning: Prior-CEA. We write it in complete form as

$$Lex.\ Min_{(E_0, \boldsymbol{E_m})} \left\{ \begin{matrix} \Theta[\varepsilon_m[P(E_0, E_m, p_m)] - P^*]\varepsilon_m[P(E_0, E_m, p_m)], \\ \varepsilon_m[C(E_0, E_m)] \end{matrix} \right\}. \tag{7}$$

Prior-CEA places the threshold P^* on the prior exceedance probability $\varepsilon_m[P(E_0, E_m, p_m)]$. As long as the exceedance probability before learning is below P^*, expected mitigation cost over all learning scenarios are minimized. Otherwise, the smallest possible prior exceedance probability is chosen.

The major problem of Prior-CEA is that even after learning, the decision maker continues to minimize mitigation cost subject to the prior constraint. She can increase the posterior exceedance probability as much as she likes as long as this is balanced by low exceedance probabilities in counterfactual learning scenarios that could have occurred in the past but eventually did not realize. This violation of consequentialism can, ultimately, lead to a "sacrifice of the climate" in some bad (high climate sensitivity) learning scenarios (Schmidt et al. 2009). In that case, second-period emissions are increased to a business-as-usual level to obtain zero mitigation cost and the actual high posterior exceedance probability can be balanced by low exceedance probabilities from the counterfactual learning scenarios.

Posterior-Prior-CEA:

Posterior-Prior-CEA (PP-CEA) determines first-period decisions according to Prior-CEA and second-period emissions according to Posterior-CEA. Unlike the other decision criteria which are intertemporal optimizations, this criterion is time-recursive. First, optimal second-period decisions $\boldsymbol{E_m^*}(E_0)$ for a given first-period decision E_0 are determined. Then, the optimal first-period decision E_0 is determined given the second-period optimum $\boldsymbol{E_m^*}(E_0)$. PP-CEA reads

$$\boldsymbol{E}_{\boldsymbol{m}}^{*}(E_0) = Lex.\ argmin_{\boldsymbol{E_m}} \begin{Bmatrix} \varepsilon_m[\Theta[P(E_0, E_m, p_m) - P^*]\, P(E_0, E_m, p_m)], \\ \varepsilon_m[C(E_0, E_m)] \end{Bmatrix}, \tag{8}$$

$$E_0^{*} = Lex.\ argmin_{E_0} \begin{Bmatrix} \Theta[\varepsilon_m[P(E_0, E_m^{*}(E_0), p_m)] - P^*]\ \varepsilon_m[P(E_0, E_m^{*}(E_0), p_m)], \\ \varepsilon_m[C(E_0, E_m^{*}(E_0))] \end{Bmatrix}. \tag{9}$$

The decision maker of PP-CEA minimizes mitigation cost always up to the allowed level of exceedance probability always with respect to her current probability distribution of climate sensitivity. However, when calculating the exceedance probability, she correctly anticipates that her second-period emissions will be determined in the same way but with respect to the future posterior distributions. The decision maker anticipates that in "good" learning scenarios she will increase emissions, while in "bad" learning scenarios she will reduce emissions even further.

The advantage of PP-CEA over Posterior-CEA is that the anticipation effect is not as extreme under certain conditions. The expectation over the posterior exceedance probability is the prior exceedance probability. This implies that maximum mitigation in the first period is avoided if there is at least one good learning scenario in which a business-as-usual continuation stays strictly below the posterior threshold. This allows for a bad learning scenario to transgress the posterior threshold. Staying below a posterior threshold with zero mitigation cost may occur for two reasons: First, the learning scenario is sufficiently "good", i.e. the bulk of the posterior is centered around sufficiently small climate sensitivities. Second, learning happens rather late when a large part of the transition to a low-carbon economy has already been achieved such that it would not be cost-minimizing anymore to go beyond the posterior threshold P^*.

Posterior-CEA, Prior-CEA and Posterior-Prior-CEA all are possible extensions of Probabilistic CEA to learning but each come with more or less severe downsides. Prior-CEA is not acceptable since it is non-consequentialist and can, moreover, "sacrifice" of the climate. Posterior-CEA disqualifies, too, due to its extreme anticipation effect, forcing first-period to be minimal only due to the fact that one learning scenario will transgress the probability threshold. Posterior-Prior-CEA would be a possible alternative given that there are good learning scenarios where transgressing the probability threshold does not reduce mitigation cost any further.

Summary and Conclusion: How to formulate a climate target under uncertainty and learning?

Our axiomatic review structures the debate on the formulation of climate targets under certainty, uncertainty and, eventually, under uncertainty and learning from a decision-theoretic perspective. A proponent of strong sustainability who prioritizes compliance with a climate target over saving mitigation cost needs to drop either the completeness, transitivity or continuity axiom introduced by von-Neumann and Morgenstern. We distinguish between a strict and a pragmatic-probabilistic interpretation of the climate target. The former requires holding global temperature below some critical level with certainty or the highest possible probability regardless of mitigation cost. The latter develops a target based also on considerations of economic feasibility and minimizes mitigation cost as long as some non-zero exceedance probability limit is not transgressed.

A proponent of the strict target can argue against the continuity axiom by claiming that the certainty of avoiding "intolerable damage" is of much greater value than a mere high probability. The pragmatic-probabilistic interpretation cannot make this argument since it is not clear why a probability increase at a specific non-zero exceedance probability should be disproportionately more dangerous than an increase at any other probability level. Moreover, as the pragmatic-probabilistic interpretation takes mitigation cost into account when setting the target, it may face the problem of coming up with reasoned preferences over a "tragic choice", i.e. to weigh high mitigation cost against high climate risk. Unlike the strict proponent, such decision maker has a reason to relax the completeness axiom as long as the decision criterion remains feasible because she seeks to avoid tragic choices that seem incomparable to her. While for the strict interpretation the utility function fails at the continuity axiom, for the pragmatic-probabilistic interpretation it fails at the completeness axiom.

The strict interpretation is quite consistent and can simply go with a lexicographic expected utility (EU) criterion (Blume et al. 1991) by minimizing exceedance probability first and mitigation cost second. This can be applied under learning as well. The pragmatic-probabilistic interpretation can go with Probabilistic CEA in the case of no-learning, i.e. it cost-effectively holds a maximum acceptable exceedance probability. However, its violation of continuity, we

think, is not desirable. Rather, it is a downside that goes along the development of reasoned preferences by using probabilistic threshold values to separate a "sustainable" from a "non-sustainable" zone in the option space. In practice, the exact level of the threshold may not be important. Sensitivity analysis around the probabilistic threshold with Integrated Assessment Models could show that the artificial discontinuity in preferences is empirically not an issue if mitigation costs do not decrease drastically beyond the threshold level. Nevertheless, we think that future conceptual work on strong sustainability under uncertainty should extend on the question why the continuity axiom should be dropped in the first place.

Our analysis shows moreover that, under learning, the pragmatic-probabilistic interpretation which does not accept "excessive" mitigation cost needs to comply with the completeness axiom, but has to drop the independence axiom: Completeness is necessary because deciding on tragic choices becomes inevitable once we anticipate bad learning scenarios. Moreover, if independence was satisfied in addition to completeness and transitivity, we would obtain a lexicographic EU criterion. This is compatible with the strict but not with the pragmatic-probabilistic interpretation since, under unbounded probability distributions, such criterion suggests reducing emissions as much as possible regardless of mitigation cost.

For the case of learning, we suggest the pragmatic-probabilistic position to apply a decision criterion that we call Posterior-Prior cost-effectiveness analysis (PP-CEA). The criterion is different former criteria (Schmidt et al. 2009; 2011) since it is not an intertemporal, but a time-recursive optimization. Under each state of knowledge (probability distribution), the decision maker applies Probabilistic CEA, i.e. she seeks to stays below the threshold value of exceedance probability in a cost-optimal manner. To calculate her exceedance probability, though, she anticipates the very behavior also in future learning scenarios with respect to the corresponding posterior distributions. Like other formulations of cost-effectiveness analysis under learning, PP-CEA decisions are transitive but break with continuity and independence.

Due to non-independence, the gist of PP-CEA is that (unlike CRA) it eventually increases mitigation ambition as much as possible once we get close to transgressing the critical temperature, yet without demanding this ambition to be as high from the beginning (like lexicographic EU). However, if climate sensitivity turns out to be lower than expected, the decision maker will invest less into mitigation than in the no-learning case since she is allowed to

increase the posterior exceedance probability up to the threshold level. This leads to maximum mitigation in the first period (before learning) if not all posterior exceedance probabilities can be reduced to threshold level and no low climate sensitivity scenario holds the level at zero mitigation cost.

One can argue that pragmatic-probabilistic strong sustainability is more intimately linked to the violation of independence than to the violation of continuity. Let us consider the willingness to pay for reducing one unit of exceedance probability (or climate risk in general) of the above criteria. The normative intuition of strong sustainability is that this willingness must become infinite once we get close or even above the critical temperature, i.e. in bad learning scenarios we should invest everything to contain climate change as much as still possible. Unlike the strict interpretation, this is not required in lower regimes of exceedance probability. While the continuity axiom implies that the willingness to pay is a continuous function of exceedance probability, the independence axiom implies that it is constant. Thus, the statement that we should not care too much about low exceedance probabilities, but invest everything once exceedance probabilities become close to 1 is, above all, a break with the independence axiom.

Can a climate target be formulated for the case of learning? It can, but our formulation may lead to an extreme anticipation effect, i.e. emissions need to be reduced as much as possible before learning due to a small chance of ending up in the worst-case learning scenario. Maximum mitigation before learning is avoided, though, if learning scenarios stay below the posterior probability threshold with a business-as-usual continuation, i.e. at zero mitigation cost. This happens if learning occurs late at a point when the energy system has already transformed to a low-carbon infrastructure such that it would not be cost-optimal in low climate sensitivity scenarios to emit more than the probabilistic target level allows. Under which specific assumptions this extreme anticipation effect is avoided is a question to be left to future research that implements PP-CEA into an empirically founded Integrated Assessment Model.

We conducted a comprehensive and systematic but not exhaustive review of possible decision criteria for strong sustainability under learning. Other formulations of lexicographic non-independent decision criteria are possible and translating the demands of strong sustainability into further axiomatic restrictions would be necessary to obtain an exhaustive picture. If the Posterior-Prior criterion leads to the trivial result of maximum mitigation, either another

lexicographic non-independent formulation is found or the axiomatic problems of EU are neglected and cost-benefit analysis (CBA) or cost-risk analysis (CRA) are used to analyze learning. CBA is based on empirically detailed and comprehensive but less transparent assessments of climate damages and may also suggest maximum mitigation in the light of fat-tailed climate uncertainty (Weitzman 2009). CRA can be calibrated to a probabilistic target and is a relatively simple and transparent target-based criterion that may be easily adjusted in the decision-making analysis. Held (2019) explains the characteristics of CRA in detail and reviews its advantages and disadvantages relative to cost-effectiveness analysis and cost-benefit analysis.

We find our method of discussing the von-Neumann-Morgenstern axioms against the background of the climate problem helpful to structure the debate on strong sustainability under uncertainty and learning. However, we also see limitations to this perspective: Not all normatively relevant aspects can be covered by axiomatics. For example, neither the differences between CBA and CRA nor the question of whether or not bounded rationality should play a role in the development of preferences can be captured by the four axioms we discussed. An axiomatic discussion is only helpful whenever two normative positions clearly understand and disagree on an axiom. It highlights their differences and assigns different classes of compatible decision criteria to them. With our contribution, we would like to encourage further research on the decision-theoretic consistency of normative positions in the sustainability discourse.

References

Ackerman, F. et al., 2009. Limitations of integrated assessment models of climate change. *Climatic Change*, 95(3–4), pp.297–315.

Allais, A.M., 1953. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, 21(4), pp.503–546.

Allen, M.R. et al., 2009. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, 458(7242), pp.1163–1166.

Baumgärtner, S. & Quaas, M.F., 2009. Ecological-economic viability as a criterion of strong sustainability under uncertainty. *Ecological Economics*, 68(7).

Blume, L., Brandenburger, A. & Dekel, E., 1991. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59(1), pp.61–79.

Bradley, R. & Stefansson, H.O., 2016. Counterfactual Desirability. *British Journal for the Philosophy of Science*, 0, pp.1–49.

Charlesworth, M. & Okereke, C., 2010. Policy responses to rapid climate change: An epistemological critique of dominant approaches. *Global Environmental Change*, 20(1), pp.121–129.

Ciriacy-Wantrup, S., 1952. *Resource Conservation. Economics and Policies*, Berkley, USA: University of California Press.

Daly, H., 2007. *Ecological Economics and Sustainable Development, Selected Essays of Herman Daly*, Northampton, USA: Edward Elgar.

Daly, H.E., 1974. The Economics of the Steady State. *The American Economic Review*, 64(2), pp.15–21.

Edenhofer, O. & Lessmann, K., 2007. Vom Preis des Klimaschutzes und vom Wert der Erde. Technischer Fortschritt und das Konzept "starker Nachhaltigkeit." In *Jahrbuch Ökologische Ökonomik. Soziale Nachhaltigkeit*. Marburg: Metropolis.

den Elzen, M.G.J. & Van Vuuren, D.P., 2007. Peaking profiles for achieving long-term temperature targets with more likelihood at lower costs. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46), pp.17931–17936.

Georgescu-Roegen, N., 1975. Energy and Economic Myths. *Southern Economic Journal*, 41(3), pp.347–381.

Gilboa, I., 2009. *Theory of Decision under Uncertainty*, New York: Cambridge University Press.

Gollier, C., 2001. *The Economics of Risk and Time*, Massachusetts Institute of Technology.

Hartwick, J.M., 1977. Intergenerational Equity and the Investing of Rents from Exhaustible Resources. *The American Economic Review*, 67(5), pp.972–974.

Held, H., 2019. Cost Risk Analysis: Dynamically Consistent Decision-Making under Climate Targets. *Environmental and Resource Economics*, 72(1), pp.247–261.

Held, H. et al., 2009. Efficient climate policies under technology and climate uncertainty. *Energy Economics*, 31, pp.S50–S61.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In T. F. Stocker et al., eds. Cambridge University Press.

IPCC, 2014. Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In O. Edenhofer et al., eds. Cambridge University Press, p. 1454.

IPCC, 2018. *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change,* V. Masson-Delmotte et al., eds., Geneva, Switzerland: World Meteorological Organization.

Kahneman, D. & Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), pp.263–292.

Loomes, G. & Sugden, R., 1982. Regret Theory: an Alternative Theory of Rational Choice Under Uncertainty. *Economic journal*, 92(368), pp.805–824.

Machina, M.J., 1989. Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty. *Journal of Economic Literature*, 27(4).

Mandler, M., 2005. Incomplete preferences and rational intransitivity of choice. *Games and Economic Behavior*, 50(2), pp.255–277.

Neubersch, D., Held, H. & Otto, A., 2014. Operationalizing climate targets under learning: An application of cost-risk analysis. *Climatic Change*, 126(3–4), pp.305–318.

Von Neumann, J. & Morgenstern, O., 1944. Theory of Games and Economic Behavior. *Princeton University Press*, p.625.

Neumayer, E., 2013. *Weak versus Strong Sustainability. Exploring the Limits of Two Opposing Paradigms.* 4th ed., Northampton, USA: Edward Elgar.

Nordhaus, W., 2013. *The Climate Casino: Risk, Uncertainty and Economics for a Warming World*, Yale University Press.

Nordhaus, W.D., 2008. *A Question of Balance: Weighing the Options on Global Warming Policies*, New Haven, London: Yale University Press.

Perman, R. et al., 2003. *Natural Resource and Environmental Economics* 3rd ed., Edinburgh Gate, UK: Pearson.

Petschel-Held, G. et al., 1999. The tolerable windows approach: theoretical and methodological foundations. *Climatic Change*, 41, pp.303–331.

Pindyck, R.S., 2013. The climate policy dilemma. *Review of Environmental Economics and Policy*, 7(2), pp.219–237.

Rockström, J. et al., 2009. Planetary Boundaries: Exploring the Safe Operating Space for Humanity. *Ecology and Society*, 14(2).

Roshan, E., Khabbazan, M.M. & Held, H., 2018. Cost-Risk Trade-Off of Mitigation and Solar

Geoengineering: Considering Regional Disparities Under Probabilistic Climate Sensitivity. *Environmental and Resource Economics*, 1–17.

Roth, R., Neubersch, D. & Held, H., 2015. Evaluating Delayed Climate Policy by Cost-Risk Analysis. *EAERE Conference Paper*.

Schellnhuber, H.J., 1998. The Scope of the Challenge. In *Earth System Analysis*. Springer, pp. 3–195.

Schmidt, M.G.W. et al., 2009. Climate Targets in an Uncertain World. *PIK Working Paper*.

Schmidt, M.G.W. et al., 2011. Climate targets under uncertainty: Challenges and remedies. *Climatic Change*, 104(3–4), pp.783–791.

Solow, R.M., 1974. Intergenerational Equity and Exhaustable Resources. *The Review of Economic Studies*, 41(1974), pp.29–45.

Tol, R.S.J., 2002. Estimates of the damage costs of climate change: Part II. Dynamic estimates. *Environmental and Resource Economics*, 21(2), pp.135–160.

Tol, R.S.J., 2009. The Economie Effects of Climate Change. *The Journal of Economic Perspectives*, 23(2), pp.29–51.

UNFCCC, 2015. *Conference of the Parties. Twenty-first session Paris, 30 November to 11 December 2015. Adoption of the Paris Agreement*,

UNFCCC, 2012. *Report of the Conference of the Parties on its seventeenth session, held in Durban from 28 November to 11 December 2011*,

UNFCCC, 2011. *Report of the Conference of the Parties on its sixteenth session, held in Cancun from 29 November to 10 December 2010. Addendum. Part Two: Action taken by the Conference of the Parties at its sixteenth session.*,

UNFCCC, 1992. *United Nations Framework Convention on Climate Change*,

Wakker, P., 1999. Justifying Bayesianism by Dynamic Decision Principles. *Working paper, Medical Decision Making Unit, Leiden University Medical Center, The Netherlands.*

Wakker, P., 1988. Nonexpected utility as aversion of information. *Journal of Behavioral Decision Making*, I(July 1987), pp.169–175.

WBGU, 2014. *Human Progress Within Planetary Guard Rails. A Contribution to the SDG Debate*,

WBGU, 1995. *Scenario for the derivation of global CO2 reduction targets and implementation strategies. Statement on the occasion of the First Conference of the Parties to the Framework Convention on Climate Change in Berlin*,

WBGU, 2011. *World in Transition. A Social Contract for Sustainability*,

Webster, M., Jakobovits, L. & Norton, J., 2008. Learning about climate change and implications for near-term policy. *Climatic Change*, 89(1–2), pp.67–85.

Weitzman, M.L., 2009. Additive Damages, Fat-Tailed Climate Dynamics, and Uncertain Discounting. *SSRN Electronic Journal*, 3, pp.1–24.