Value Trade-Offs and Sustainability Policy: An Economic Analysis

Von der Fakultät Wirtschaftswissenschaften
der Leuphana Universität Lüneburg

zur Erlangung des Grades
Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)
genehmigte

Dissertation

von

Nikolai Hoberg

aus Stuttgart

Eingereicht am: 13. September 2013

Mündliche Prüfung: 18. Dezember 2013

Erstgutachter: Prof. Dr. Stefan Baumgärtner

Zweitgutachter: Prof. Dr. Michael Schefczyk

Prüfungsausschuss: Prof. Dr. Stefan Baumgärtner, Vors.
Prof. Dr. Michael Schefczyk
Prof. Dr. Thomas Wein

Die einzelnen Beiträge des kumulativem Dissertationsvorhabens sind oder werden wie folgt in Zeitschriften veröffentlicht:

Hoberg, N. and Baumgärtner, S. (2013). Value Pluralism, trade-offs and efficiencies. Submission planned for 2014.

Baumgärtner, S., S. Glotzbach, N. Hoberg, M.F. Quaas and Stumpf, K.H. (2012). Economic analysis of trade-offs between justices, Intergenerational Justice Review, 1/2012, p. 4-9.

Hoberg, N. and Baumgärtner, S. (2011). Irreversibility, ignorance and the intergenerational equity-efficiency trade-off, University of Lüneburg Working Paper Series in Economics, No. 198, February 2011. (revised on February 24, 2012)

Hoberg, N. (2013). Merit good arguments in sustainability discussions: Challenges and opportunities. Submission planned for 2014.

Elektronische Veröffentlichung des gesamten kumulativen Dissertationsvorhabens inklusive einer Zusammenfassung unter dem Titel:

Value Trade-Offs and Sustainability Policy: An Economic Analysis

Veröffentlichungsjahr: 2014

Veröffentlicht im Onlineangebot der Universitätsbibliothek unter der URL:
www.leuphana.de/ub

Contents

List of Figures

Acknowledgements

First, I would like to thank Prof. Dr. Stefan Baumgärtner for his intellectual support and guidance in all the steps that lead up to this dissertation. His intellectual openness, rigor and patience have first drawn me to the study of economics, and later allowed me to take my own steps in research of which one result is this dissertation. Further, I would like to thank Prof. Dr. Michael Schefczyk and Prof. Dr. Thomas Wein for their support and comments that helped me to stay orientated during this dissertation project.

This dissertation has profited from the serious intellectual discussions and numerous humorous exchanges with all of my colleagues at the Institute of Economics and the Sustainability Economics Group, among others with David Abson, John-Oliver Engler, Joachim Fünfgelt, Stefanie Glotzbach, Lutz Göhring, Roland Olbrich, Sebastian Strunz and Klara Stumpf. Thanks for the good times, fond memories and for becoming friends. I also would like to thank the scientific community at large for comments and feedback on the individual papers at international conferences, seminars and in the review process for peer-reviewed journals.

Finally, I am especially grateful to my family Rolf, Ilona, Wahlburg, and Barbara, and my friends for their encouragement and help on all different levels. I am profoundly grateful to Anna for being there for and with me, and for showing me the value in happiness.

Introduction

Efficiency in the allocation of resources is often heralded as a central normative argument in favor of (free) markets. Increases in economic development, life expectancy, literacy rates, and reductions in child mortality are in turn attributed to the increased use of markets to allocate resources (World Bank 2012). Yet, there are increasing signs of global environmental damages that raise concerns if there are other normative objectives or values at stake even with regard these impressive advances in current human well-being (e.g. Dasgupta 2010). Those concerned with sustainability point to examples such as climate change, biodiversity loss, depletion of fishstocks, and loss of arable land to highlight the intergenerational effects of the current development (MEA 2005, IPCC 2007, UNEP 2012). Thus, the evaluation of these diverse developments requires a concept of sustainability that is able to balance different values and potential trade-offs between values. For example, the influential Brundlandt definition balances future and current well-being by defining sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (WCED 1987: 43).

Economics as a science provides powerful methods to analyze such trade-offs between values. Probably the most famous value trade-off in economics is the trade-off where, in contrast to the second welfare theorem, there is a conflict between equity and efficiency (Putterman et al. 1998). In this vein, some have studied the trade-off between sustainability and efficiency (e.g. Krysiak 2009). Of course, there are other value trade-offs that refer more directly to current and future well-being such as the discussion on the modeling of and ethical justification for the 'right' discount rate to evaluate the costs and benefits of greenhouse gas emissions (e.g. Weitzman 2007, Roemer 2011, Dasgupta

2011). Generally, this discussion of value trade-offs in economics requires reflection of the normative foundations of economic theory and a methodological approach fit to this challenge. This ethical reflection is ongoing (e.g. Sen 1977, Sen 1979, Hausman and McPherson 2006) and poses specific problems for the economic analysis of sustainability problems. In this sense, 'sustainability economics' aims at a reflection of such topics as the role of nature, and future generations in economics as well as different conceptions of justice in economic theory (Baumgärtner and Quaas 2010).

In this thesis I discuss trade-offs between values in different economic models and combine these models with philosophical reasoning so as to ensure rigorous results that can be clearly related to the philosophical literature. In the economic literature there are many results on particular relationships between particular values such as equity and efficiency or inter- and intragenerational well-being. This leads to the first question I take up in this thesis: Can a general definition of efficiency with respect to values provide a unifying interpretation of the diverse results on relationships between particular values? The second question I take up is more specific and considers sustainability and relationships between values: What can economic analysis contribute to the discussion of value trade-offs if sustainability is defined in terms of inter- and intragenerational justice? The third question I take up focuses on the emergence of value trade-offs due to the characteristics of the intertemporal setting of sustainability problems: Does intertemporal policy-making pose conditions that lead to a unique mechanism for a trade-off between sustainability and Pareto-efficiency? The fourth question I take up starts with the observation that many economic analyses of trade-offs are based on a conception of well-being, i.e. an idea on what makes individual better-off. This leads to the fourth question: How do different conceptions of individual well-being allow to make value trade-offs in sustainability discussions?

By answering these questions this thesis contributes to inter- and transdisciplinary research on sustainability problems as in Baumgärtner et al. (2008) on two levels of abstraction: (i) it reflects concepts of economics in light of philosophical theories, for example, by discussing how the concept of efficiency and the philosophical discussion on values can be combined in economic theory; and (ii) it uses generic models to derive insights into relationships between values, for example, by studying unique mechanisms for trade-offs between values in intertemporal settings. This thesis does not comprise empirical studies on concrete case studies, as the third level of abstraction in Baumgärtner et al. (2008), but contributes to the refinement of the conceptual basis on which empirical research builds.

In the following I will first briefly introduce all the four papers in this thesis, summarize their results, and discuss their limitations, possible extensions and draw a short conclusion on the general contribution of this thesis.

Research Papers

This thesis comprises four research papers whose relations are illustrated in Figure 1. Paper 1 provides a general philosophical and economic framework for the discussion of value efficiency and relationships between values. The following two papers discuss specific value trade-offs in different contexts. Paper 2 directly applies the ideas from the first paper to the concept of sustainability and discusses the relationship between intra- and intergenerational justice. Paper 3 discusses how irreversibility and ignorance cause a trade-off between equity and efficiency in an intergenerational setting. Paper 4 discusses conceptions of well-being that underlie merit good arguments and how these conceptions impact the ability to make comparisons of well-being.

Paper 1: Value pluralism, trade-offs and efficiencies

The paper "*Value pluralism, trade-offs and efficiencies*" provides a general philosophical and economic framework for the analysis of relationships between values. In economics, there are many examples of relationships between values such as the equity-efficiency trade-off (Putterman et al. 1998) or the win-win relationship between income equality and social mobility (Corak 2013). These results on relationships between values do not try to establish which combinations of values are desirable, but rather show what combinations of values are feasible in a given context, i.e. they are concerned with the shape of the set of feasible states of affairs in terms of values (e.g. Le Grand 1990, Cowen

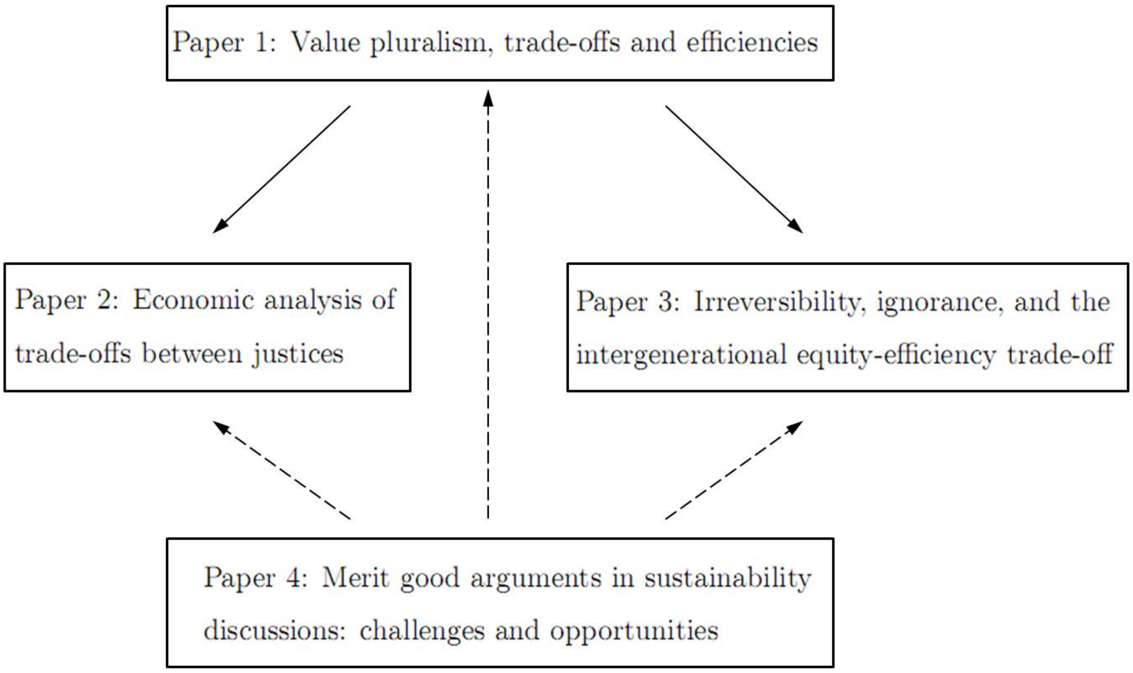
Paper 1: Value pluralism, trade-offs and efficiencies
Paper 2: Economic analysis of trade-offs between justices
Paper 3: Irreversibility, ignorance, and the intergenerational equity-efficiency trade-off
Paper 4: Merit good arguments in sustainability discussions: challenges and opportunities

Figure 1: Illustration of the relations between the research papers: Paper 1 discusses value trade-offs and value efficiency which provides general insight into relationships between values. Paper 2 and Paper 3 discuss trade-offs between specific values: equity and efficiency as well as intra- and intergenerational justice. Paper 4 reflects on merit good arguments and the related conceptions of well-being which are relevant to any situation where values are based on utility.

2007). This practice follows authors such as Dasgupta (2005) in employing economics as a method to discuss the implications of different ethical theories rather than establishing what values there should be.

In order to provide a unifying interpretation of the numerous results on relationships between particular values in particular contexts, we introduce the criterion of value efficiency. This is an efficiency criterion with respect to normative objectives. It is a rather uncontroversial criterion as it does not require weighing the degree of attainment of different values, but rather defines a state of affairs as value efficient if it is impossible to increase the degree of attainment of one value without necessarily reducing the degree of attainment of any other value. While this notion of efficiency has been more or less explicitly discussed in the literature before (e.g. Sen 1985, Le Grand 1990, Dasgupta 2005, Pattanaik and Xu 2012), we relate it explicitly to different sets of feasible states of affairs to provide a unifying interpretation of specific results on relationships between values. Regarding the philosophical background, we build on the philosophical discussion on value monism and value pluralism. Advocates of value monism argue that all values can be reduced to a single encompassing intrinsic value – such as wellbeing or happiness (e.g. Bentham 1988 [1776]). In contrast, advocates of value pluralism argue that there are multiple intrinsic values – such as equality and liberty (e.g. Berlin 1969). By engaging in this philosophical debate we seek to ensure not only rigorous results, but also results that can be interpreted meaningfully. It makes a difference whether there is one or many intrinsic values when interpreting trade-offs, for example, between the value of agricultural production and the value of biodiversity. In the paper, we build a small economic model to incorporate values as binary relations into an economic framework and go on to define value efficiency as efficiency with respect to values. In the next step, we distinguish three different relationships between values in our model: trade-off, win-win and independence relationships. Depending on the employed ethical theory, these relationships can be between instrumental values, intrinsic values or both. In the final step, we take the discussion of relationships between values to the familiar individualistic framework of welfare economics where individual preferences play a central role in determining values.

The paper shows how relationships must be carefully conceptualized so as to yield normatively meaningful interpretations. For example, relationships between an instrumental and an intrinsic value are difficult to interpret if the intrinsic value behind the instrumental value is not clearly explicated – as for example in the case of the equity-efficiency trade-off (Le Grand 1990). Regarding the connection between value efficiency and the different relationships between values, we show that there can be independence and win-win relationships in a value-efficient state of affairs, if there are three or more values. For example, it can be when there are three values that in a value-efficient state the higher attainment of one value necessarily leads to a higher attainment of a second value (win-win relationship between first and second value) as long as it also necessarily reduces the attainment of a third value (trade-off relationship between first and third value). Further, we show that in a Pareto-efficient state there can be a any relationship between non-preference values which are values that are not reducible to individual preferences. For example, equality and liberty (as in Sen 1970) are non-preference values if people have self-regarding preferences, so that in a Pareto-efficient state there can be a win-win relationship between these two non-preference values.

The very general approach to values in this paper has at least two limitations. One limitation lies in the fact that it does not employ values derived from a concrete ethical theory. This leads to results which hold generally, but which do not provide insights into specific ethical problems. As we wanted to characterize the connection between value efficiency and relationships between values in a general manner, the contribution of the paper must be seen in this regard. This limitation is taken up in Paper 2 where decision-making for sustainability, defined in terms of intra- and intergenerational justice, is studied as a specific ethical problem. A second limitation lies in the fact that we do not provide a concrete economic model to elucidate the assumptions on individual preferences, institutions and instruments necessary to generate all the different relationships on a feasible set with three values. As our results hold for value-efficiency in general, they hold for any concrete feasible set. A fruitful extension of this paper would be the study of relationships between values in a concrete economic model with particular values that are derived from a concrete ethical theory. This limitation is taken up in Paper 3 where the effect of irreversibility and ignorance on intergenerational equity and Pareto-efficiency are studied in a concrete economic model.

Paper 2: Economic analysis of trade-offs between justices

The paper "*Economic analysis of trade-offs between justices*" applies the general discussion of value trade-offs to the specific discussion on the relationship between intra- and intergenerational justice in the sustainability concept. In economics the conceptions of justice are commonly treated with the tool of a social welfare function. However, these mainly represent different conceptions of distributive justice and take utility as their measure of individual advantage (e.g. Sen 2008). At least since the publication of John Rawls 'Theory of Justice' in (1971) there has been increased interest in the question on how to incorporate more complex conceptions of justice in economics. For example, there are contributions in economic theory that discuss different measures of individual advantage such as capabilities or Rawls primary goods (e.g. Roemer 1996), or try to incorporate procedural concerns into their analysis (e.g. Sen 1997). In this vein, we seek to combine the philosophical discussion on conceptions of justice with the economic discussion on how to represent 'justice' in economic theory for the case of sustainability.

In our conceptual discussion, we define sustainability in terms of intra- and intergenerational justice. We start from three basic assumptions: (i) we assume that both justices are values of equal importance so as to ensure meaningful relationships between the two, (ii) that it is possible to measure the degree of attainment of these justices, and (iii) that it is possible to describe the outcomes of using instruments of justice in a given context. We bring in economic analysis by considering scarcity in the use of instruments of justice which follows Robbins definition of economics as the science that "studies human behaviour as a relationship between ends and scarce means which have alternative uses" (Robbins 1932: 15). We then go on to discuss the contribution of economic analysis to this discussion of justices by defining efficiency in the use of instruments of justice. This criterion of efficiency says that a situation is 'efficient' with regard to the two justices if it is impossible to improve the attainment of one justice without decreasing the attainment of another one. We proceed to define different relationships between these two values, as in Paper 1, and consider their connection to efficiency in this particular context.

The paper shows how economic analysis can be employed to analyzing relationships between intra- and intergenerational justice. The contribution of this paper, also in light of the first limitation of Paper 1, lies in its clear application of economic concepts to concrete ethical theories, namely, conceptions of intra- and intergenerational justice in the sustainability context. Specifically, we discuss how economics can help in discussions of justice by delineating an opportunity set in terms of the two justices. Further, we show that with these two values there is a trade-off between intra- and intergenerational justice when the instruments of justice are used efficiently. Conversely, a win-win relationship signals that instruments of justice are used inefficiently. Economic analysis of the opportunity set also allows the discussion of opportunity costs of attaining one justice to a higher degree in terms of the reduction in degree of attainment of the other – which is highly useful information for societal decision-making. For example, sustainability policy might be done differently, if it could be clearly shown that a small burden on the current generation leads to a large increase in future well-being.

The rather conceptual approach to the discussion of relationships between intra- and intergenerational justice in the paper has its limitations. For example, one limitation lies, as with Paper 1, in the fact that it cannot provide the specific assumptions necessary for the existence of different relationships between the two justices or the specific information on technology and institutions necessary to describe the outcomes of using instruments of justice. The paper does not show analytically how the interactions of individuals and different policy instruments leads to the emergence of different relationships. While the paper provides a clear conceptual approach to such problems, it would be interesting to see how, for example, a win-win relationship between intra- and intergenerational justice actually emerges in an economic model with multiple agents for a concrete conception of justice. A second limitation of the results on efficiency and trade-offs in this paper derives from the fact that it employs only two values. As seen in Paper 1, it makes a difference to the connection of trade-off and efficiency if there are three or more values. This means that if there are values in sustainability policy apart from intra- and intergenerational justice that are considered important, then the result that a win-win relationship between intra- and intergenerational justice signals inefficiency requires reinterpretation.

Paper 3: Irreversibility, ignorance and the intergenerational equity-efficiency trade-off

The paper "*Irreversibility, ignorance and the intergenerational equity-efficiency trade-off*" discusses another specific trade-off, namely the trade-off between Pareto-efficiency and intergenerational equity.[1] Following the second fundamental theorem of welfare economics equity and Pareto-efficiency are usually treated separately. Yet, there are many results on mechanisms that lead to an intragenerational equity-efficiency trade-off such as the inefficiency due to incentive distortions from personal income taxation (Putterman et al. 1998). In the spirit of the second fundamental theorem of welfare economics, Howarth and Norgaard (1990, 1992) show how Pareto-efficiency and intergenerational equity are simultaneously attainable in an overlapping generations model. However, there are only a few studies on mechanisms that lead to intergenerational equity-efficiency trade-offs. For example, Krautkraemer and Batina (1999) show how a non-decreasing utility constraint in a model with a renewable resource can lead to Pareto-inefficiency. We focus on irreversibility and ignorance, as a strong form of uncertainty, as two central characteristics of intergenerational problems. The effects of irreversibility and uncertainty have been widely studied in environmental and resource economics (Arrow and Fisher 1974, Henry 1974). Still, there are only few studies that focus on ignorance or unawareness and thereby incorporate situations where there are unforeseen contingencies in economic models (e.g. Dekel et al. 1998).

In this paper, we are the first to include unawareness about an intergenerational externality in an environmental economic setting. For this, we develop a model with two non-overlapping generations who use a non-renewable resource and circulating capital in production. As an example, one can think of the case of climate justice between historic and future emitters. Here, the first generation corresponds to historic emitters who use fossil fuels under unawareness of the effect of greenhouse gas emissions on future generations. The second generation corresponds to future emitters who then suffer the effect of climate change and receive less of the fossil fuels for production. We introduce the two normative objectives (or values) of Pareto-efficiency and intergenerational equity to this model through a social planner who can use an intergenerational capital transfer and a restriction on resource use to achieve these goals.

The paper shows a genuine intertemporal mechanism for a trade-off between intergenerational equity and Pareto-efficiency due to two characteristics specific to intertemporal

[1] In this introduction equity and justice are used interchangeably.

decision-making: irreversibility in resource use and unawareness about an intergenerational externality. The contribution of this paper, in light of the first limitation of Paper 1 and the limitations of Paper 2, lies in the fact that the results are derived for concrete values, intergenerational equity and Pareto-efficiency, in a concrete economic model where assumptions on technology, instruments and incentives are clearly explicated. While unawareness about the effects of actions in human-environment systems is pervasive, we are the first to include this strong form of uncertainty into an intergenerational setting. For the case of climate justice, we infer that current climate policy faces a trade-off between an equitable or an efficient policy, as such policy is enacted after fossil fuels have been irreversibly used under unawareness about the effect of greenhouse gas emissions. More generally, we find that one should beware of irreversibility in case of unawareness as it reduces ones ability to react to unforeseen negative (or positive) effects. Further, regarding unawareness itself, we find that one should try to reduce the potential costs of unawareness, for example, by investing in research. Regarding the values of Pareto-efficiency and equity, we find that weaker normative objectives might be less prone to unawareness. For example, by only aiming at the satisfaction of future basic needs one requires less information about technology and future preferences which leads to a lesser likelihood of problems originating in unawareness.

The economic model we use in this paper is limited. For example, it does not consider infinitely many generations, overlapping generations, or other sources of unawareness. In the discussion section of the paper we discuss these and other extensions of the model and show that these do not change the conclusions we draw from our more simple model. Still, it would be interesting to see the effects of irreversibility and unawareness in more encompassing general equilibrium models with several agents per generation. In a more comprehensive model one could put more emphasis on how unawareness is actually resolved and how this newly acquired information travels between agents. For example, one could study the case where some groups adapt to new information due to research more quickly than others. Here, it would also be interesting to see how this affects the intragenerational distribution between agents. Another interesting extension would be to consider different kinds of unawareness which can be reducible or not reducible.

Paper 4: Merit good arguments in sustainability discussions: Challenges and Opportunities

The paper "*Merit good arguments in sustainability discussions: Challenges and Opportunities*" discusses different conceptions of well-being behind merit good arguments and how these impact the ability to make trade-offs between the well-being of generations. The concept of merit goods was defined by Musgrave (1957, 1959) to refer to situations "where evaluation of a good (its merit or demerit) derives not simply from the norm of consumer sovereignty but involves an alternative norm" (Musgrave 1987: 579). That is, merit good arguments justify government intervention in markets differently from externalities or distributive concerns as they deviate from individual preferences for the case of a merit good. This allows taxes and subsidies on merit goods even in a situation which would be considered Pareto-efficient and equitable in the absence of merit good arguments. This challenges the common assumption in economics that every free action by individuals is also good for their well-being (Hausman and McPherson 1993: 680). This challenge is central to value trade-offs in economics as the values involved are mostly defined on the basis of individual utilities. For example, the common equity-efficiency trade-off often relies utility-based criteria such as Pareto-efficiency and utility-based equity criteria.

In the paper, I use Besley's (1988) simple model for merit good arguments to clarify the concept of merit goods analytically by distinguishing utility functions from merit utility functions that deviate from individual preferences for a specific merit good. In a next step, I follow Goodin's (1989) discussion of merit good arguments to relate justifications for the deviation from individual preferences to different conceptions of well-being. Finally, I consider what challenges and opportunities merit good arguments raise in discussions of sustainability problems.

The paper shows how merit good arguments rely on specific conceptions of well-being and how these create several challenges and opportunities in sustainability discussions. In relation to the other papers, the contribution of this paper lies in its reflection of different conceptions of well-being. For example, this relevant in the debate on value-efficiency vs. Pareto-efficiency in Paper 1. Following the analytical model in Besley (1988) and the discussion in Goodin (1989), I relate merit good arguments and the implied merit utility functions to two conceptions of well-being. Each of these conceptions justifies paternalist policies in a different manner. First, an informed preference satisfaction conception of well-being justifies the temporary deviation from individual preferences when people are not fully informed on the importance of certain goods – what Goodin (1989) calls the 'retrospective rationality' justification. Second, a perfectionist conception of well-being says that some goods are important irrespective of their contribution to the satisfaction of individual preferences which Goodin (1989) connects with the 'political preference' justification. This is the idea that people express different preferences in the market place than in the political arena, so that democratic political processes can justify policies that go against individual 'market preferences'. In a final step, I outline what challenges and opportunities these conceptions of well-being raise in sustainability discussions. For example, when a future representative individual is used to evaluate the effects of some intergenerational policy such as climate policy and one deviates from given future preferences, then this presupposes that future democratic outcomes justify this deviation. This shows the challenge that merit good arguments have higher information requirements in an intergenerational setting than in an intragenerational setting where democratic outcomes are more easily observable. Another example, is the opportunity to study merit good arguments in situations where there is a negative externality originating from a merit good. Here, different normative objectives might interact: a paternalist tax on the merit good could interact with a Pigouvian tax that is introduced to achieve Pareto-efficiency. I conclude that while merit good arguments are currently not discussed as much as other conceptions of well-being, such as the capability approach by Sen (1980), they provide fruitful analytical results for situations where individual actions do not always improve individual well-being.

One limitation of this paper lies in the lack of a concrete economic model to explicate the consequences of merit good arguments in an intergenerational setting. Still, due to the clarification provided in the paper, this could be easily done by adopting a small model with two generations and a social planner who pursues Pareto-efficiency and a sustainability criterion based on a merit utility function. This would also allow one to analyze policies and equilibria with respect to different criteria for Pareto-efficiency, one with respect to given utility functions, and one with respect to merit utility functions.

Conclusion

Taking a step back from the individual papers one can see that these illustrate a general theme regarding the combination of economic analysis and philosophical reasoning. This is the "trade-off between simplicity, generality, and theoretical precision on the one hand, and plausibility and recognition of complexity and "messiness" on the other" (Hausman and McPherson 1993: 679). This trade-off opens two dimensions for interpreting the results and relations between the individual papers, as illustrated in Figure 1, in this thesis: One concerns the level of detail of economic models regarding assumptions on individual preferences, instruments, physical and ecological parameters. Here, for example, a generic model of a renewable resource would stand on one extreme, whereas a detailed integrated assessment model for a climate system would stand on the other. The other concerns the level of detail of philosophical theory regarding conceptions of well-being, theories of justice and political philosophy. Here, for example, general questions on theories of justice would stand on one extreme, and specific ethical problems related to discounted utilitarianism on the other.

The individual papers in this thesis balance this trade-off in different ways by studying sustainability at different levels of detail. Paper 1 "Value pluralism, trade-offs and efficiencies" develops the general criterion of value efficiency and employs this criterion in studying different relationships between values with minimal assumptions regarding the set of feasible states of affairs. This analysis is limited in two ways: (i) it does not employ a specific ethical theory that determines specific values and (ii) it does not provide a concrete economic model where instruments, technology and individual preferences are explicated. Paper 2 "Economic analysis of trade-offs between justices" addresses limitation (i) as it concerns specific ethical theories by employing a definition of sustainability in terms of intra- and intergenerational justice. Paper 3 "Irreversibility, ignorance, and the intergenerational equity-efficiency trade-off" addresses limitation (ii) as it shows the trade-off between intergenerational equity and Pareto-efficiency in a concrete economic model where irreversibility and ignorance put uniquely intertemporal constraints on feasible policies. Paper 4 "Merit good arguments in sustainability discussions: Challenges and opportunities" takes a step back from concrete relationships between values as it addresses specific conceptions of well-being that underlie merit good arguments. This discussion of different conceptions of well-being is the ethically most detailed analysis in the thesis and is relevant to any situation where values are defined on the basis of utility.

Still, in view of the generality-complexity trade-off, the contribution of this thesis lies rather on the general and theoretical side of the economic and philosophical discussion on values in sustainability policy. Hausman and McPherson argue that highlighting the complexity of actual moral judgements can help economic analysis by "discouraging economists from premature or overly sweeping generalization" (Hausman and McPherson 1993: 679). One may hope that if this thesis elucidates the discussion on some values in sustainability policy, that this does not incur this problem, but rather carefully spills over to the more complex and messy side of everyday decision-making for sustainability.

Bibliography

Arrow, K. and Fisher, A. (1974). Environmental preservation, uncertainty and irreversibility. *Quarterly Journal of Economics*, 88(2):312–319.

Baumgärter, S. and Quaas, M. (2010). What is sustainability economics? *Ecological Economics*, 69:445–450.

Baumgärter, S., Becker, C., Frank, K., Müller, B. and Quaas, M. (2008). Relating the philosophy and practice of ecological economics. The role of concepts, models and case studies in inter- and transdisciplinary sustainability research. *Ecological Economics*, 67:384–393.

Bentham, J. (1988). *A Fragment on Government.* Cambridge University Press, Cambridge.

Berlin, I. (1969). *Four Essays on Liberty.* Oxford University Press, Oxford.

Besley, T. (1988). A simple model for merit good arguments. *Journal of Public Economics*, 35(3):371–383.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102.

Cowen, T. (2007). The importance of defining the feasible set. *Economics and Philosophy*, 23(1):1–14.

Dasgupta, P. (2005). What do economists analyze and why: values or facts? *Economics and Philosophy*, 21(2):221–278.

Dasgupta, P. (2010). Nature's role in sustaining economic development. *Philosophical Transactions of the Royal Society B*, 365:5–11.

Dasgupta, P. (2011). The ethics of intergenerational distribution: Reply and response to John E. Roemer. *Environmental and Resource Economics*, 50:475–493.

Dekel, E., Lipman, L. and Rustichini, A. (1998). Recent development in modeling unforeseen contingencies. *European Economic Review*, 42(3):523–542.

Goodin, R. (1989). Stars to steer by: the political impact of moral values. *Journal of Public Policy*, 9(3):241–259.

Hausman, D. and McPherson, M. (1993). Taking ethics seriously: Economics and contemporary moral philosophy. *Journal of Economic Literature*, 31(2):671–731.

Hausman, D. and McPherson, M. (2006). *Economic Analysis, Moral Philosophy, and Public Policy*, 2nd edition. Cambridge University Press, Cambridge.

Henry, C. (1974). Investment decisions under uncertainty: the "Irreversibility Effect". *American Economic Review*, 64(6):1006–1012.

Howarth, R.B. and Norgaard, R.B. (1990). Intergenerational resource rights, efficiency, and social optimality. *Land Economics*, 66(1):1–11.

Howarth, R.B. and Norgaard, R.B. (1992). Environmental valuation under sustainable development. *American Economic Review – Papers and Proceedings*, 82(2):473–477.

Intergovernmental Panel on Climate Change (IPCC) (2007). *Climate Change 2007: The Physical Science Basis; Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge.

Le Grand, J. (1990). Equity versus efficiency: the elusive trade-off. *Ethics*, 100(3):554–568.

Krautkraemer, A.J. and Batina, G.R. (1999). On sustainability and intergenerational transfers with a renewable resource. *Land Economics*, 75(2):167–184.

Krysiak, F.C. (2009). Sustainability and its relation to efficiency under uncertainty. *Economic Theory*, 41(2):297–315.

Millenium Ecosystem Assessment (MEA) (2005). *Ecosystems and Human Well-Being.* Island Press, Washington, DC.

Musgrave, R. (1957). A multiple theory of budget determination. *FinanzArchiv*, New Series, 17(3):333–343.

Musgrave, R. (1959). *Theory of Public Finance.* McGraw Hill, New York.

Musgrave, R. (1987). Merit goods. In: Eatwell, J., Milgate, M. and Newman, P. (eds), *The New Palgrave: A Dictionary of Economics*, Vol. 4, Macmillan, London. 792–793.

Pattanaik, P. and Xu, Y. (2012). On dominance and context-dependence in decisions involving multiple attributes. *Economics and Philosophy*, 28(2):117-132.

Putterman, L. Roemer, J., and Silvestre, J. (1998). Does egalitarianism have a future? *Journal of Economic Literature*, 36(2):861-902.

Rawls, J. (1971). *A Theory of Justice.* Harvard University Press, Cambridge, Mass.

Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science.* Macmillan, London.

Roemer, J. (1996). *Theories of Distributive Justice.* Harvard University Press, Cambridge, Mass.

Roemer, J. (2011). The ethics of intertemporal distribution in a warming planet. *Environmental and Resource Economics*, (48):363–390.

Sen, A. (1970). The impossibility of a paretian liberal. *Journal of Political Economy*, 78(1):152–157.

Sen, A. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, 6(4):317–344.

Sen, A. (1979). Personal utilities and public judgements: or what's wrong with welfare economics. *Economic Journal*, 89:537–558.

Sen, A. (1980). Equality of what? In: McMurrin, S. (ed), *The Tanner Lecture on Human Values*, Vol. 1, Cambridge University Press, Cambridge. 197-220.

Sen, A. (1985). Well-being, agency and freedom: the Dewey Lectures 1984. *Journal of Philosophy*, 82(4):169–221.

Sen, A. (1997). Maximization and the act of choice. *Econometrica*, 65(4):745–779.

Sen, A. (2008). Justice. In: Durlauf, S., and Blume, L. (eds), *The New Palgrave: A Dictionary of Economics*, 2nd edition, Macmillan, London. 792–793.

United Nations Environment Programme (UNEP) (2012). *Global Environment Outlook 5: Environment for the Future We Want.* UNEP, Nairobi, Kenya.

Weitzman, M. (2007). A review of the Stern Review on the economics of climate change. *Journal of Economic Literature*, 45(3):703–724.

World Bank (2012). *Inclusive Green Growth: The Pathway to Sustainable Development.* World Bank, Washington, D.C.

World Commission on Environment and Development (WCED) (1987). *Our Common Future.* Oxford University Press, Oxford.

Paper 1:
Value pluralism, trade-offs and efficiencies

Value pluralism, trade-offs and efficiencies

NIKOLAI HOBERG* and STEFAN BAUMGÄRTNER

Department of Sustainability Science and Department of Economics,
Leuphana University of Lüneburg, Germany

Abstract:
Political debates often revolve around the factual relationships between normative objectives or values, such as the equity-efficiency trade-off, rather than the desirability of specific values. To provide a unifying interpretation for the numerous results on relationships between particular values, we build on the philosophical literature on value monism vs. value pluralism, and follow Sen (1979a) and others in defining *value-efficiency*, i.e. efficiency with respect to values. We show that a win-win relationship between values is no indication of value-inefficiency when there are three or more values, and that in a value-efficient state of affairs there is a trade-off between at least two values. We contrast Pareto-efficiency with value-efficiency and show that there can be a win-win relationship between values in a Pareto-efficient state of affairs.

JEL-Classification: D63, H43, H00

Keywords: efficiency, value pluralism, value trade-offs, welfarism

*Corresponding author: Sustainability Economics Group, Leuphana University of Lüneburg, P.O. Box 2440, D-21314 Lüneburg, Germany, phone: +49.4131.677-2715, fax: +49.4131.677-1381, email: hoberg@uni.leuphana.de.

1.1 Introduction

Does an increase in income equality reduce economic growth? How should society balance the objectives relating to income equality and economic growth? These are two central questions discussed under the well-known trade-off between equality and efficiency (Le Grand 1990: 557). The first is about the *feasibility* of combinations of degrees of attainment of values. The second is about the *desirability* of combinations of degrees of attainment of values. Here, some (e.g. Dasgupta 2005, Cowen 2007) argue that answering the first question is more important because merely clarifying the feasibility of actions can reduce conflict between different ethical perspectives. For this purpose, it is important to know the (factual) relationship between two values, e.g. whether there is a trade-off or a win-win relationship between income equality and economic growth.

Apart from the large literature on the equality-efficiency trade-off (Okun 1975, Putterman et al. 1998, Blank 2002), many more examples of such relationships can be found in the economics literature: such as the relationship between income inequality and intergenerational income mobility (Björklund and Jäntti 1997, Corak 2013), the relationship between income inequality and subjective happiness (Alesina et al. 2004), the relationship between economic freedom and economic growth (DeHaan et al. 2006), the relationship between poverty reduction and environmental degradation (Baland et al. 2010), or the effect of payments for environmental services on Pareto-efficiency and poverty reduction (Engel et al. 2008). The relationships between values also feature in political debates, for example, in the one on the social effects of economic growth such as the World Bank's report on green growth (World Bank 2012).

All of these results on particular relationships between particular values depend on the form of the feasible set in each context. That is, the form of the feasible set determines whether there are, for example, trade-offs or win-win relationships between values. Describing the feasible set in terms of normative objectives or values and interpreting relationships between values requires a philosophical-economic framework. Foremost, this demands a clear definition of value. For example, it must be clear why equality and efficiency are values and why the relationship between the two is normatively meaningful. In this context it must also be clear whether equality and efficiency are instrumental to some other value such as social welfare or whether they are both intrinsic values. In light of this, we build on the philosophical literature for a definition of value in ethical theories and the debate on intrinsic values (value monism vs. value pluralism).

A criterion that should be straightforward in studying the feasible set in terms of values is value-efficiency, i.e. efficiency with respect to normative objectives. It does not require weighing the degree of attainment of different values, but rather defines a state of affairs as value-efficient if it is impossible to increase the degree of attainment of one value without necessarily reducing the degree of attainment of any other value. This efficiency criterion has been more or less explicitly discussed in the philosophical and economic literature (e.g. Sen 1979a, Le Grand 1990, Raz 1997, Dasgupta 2005, Pattanaik and Xu 2012). This leads to the question on how value-efficiency connects to relationships between values. In this paper, we study this question by providing a framework that starts from the role of values in ethical theories and leads up to the discussion of the connection between value-efficiency and relationships between values. This is to provide a unifying perspective for the interpretation of the many specific results on relationships between particular values such as the examples given above on relationships between economic growth, income equality and economic freedom. In a further step, we explicitly discuss value-efficiency in the specific framework of welfare economics when there are values that are not reducible to individual preferences (sometimes called 'non-utility values'). This allows us to discuss the conflict between value-efficiency and the familiar criterion of Pareto-efficiency.

The paper is structured as follows. Section 1.2 reviews the philosophical literature and gives a definition of value. Section 1.3 introduces a microeconomic model, gives a definition of value-efficiency and a definition of different relationships between values. Section 1.4 discusses the connection between Pareto-efficiency and value-efficiency. Finally, Section 1.5 concludes.

1.2 Philosophical background

1.2.1 Definition of value

For the analysis of relationships between values one must be clear on what a *value* is and what it is not. Many definitions[1] of value focus on 'good' states of affairs. The examples from the introduction on values such as equality or efficiency also imply statements on good states of affairs, i.e. a state of affairs is better if it is more equal or more efficient. In light of this, we define a value in the spirit of Hurka (2006) and Chang (1997b) as follows:

Definition 1 (Value)
A *value* is a consideration which allows a comparison of states of affairs in terms of their goodness.

This definition refers to *states of affairs* which excludes rules and procedures from the definition of value.[2,3] A *comparison* examines the differences between different states of affairs in terms of a *consideration.* What constitutes the *goodness* of a state of affairs is determined in an ethical theory. For example, the conception of goodness in 'outcome utilitarianism' says that a state of affairs is better than another one if its sum of individual utilities is larger (Sen 1979b). Of course, there can be different conceptions of goodness, for example, there can be different equity criteria over individual utilities (Sen

[1] For Chang (1997b: 5) a value "is any consideration with respect to which a meaningful evaluative comparison can be made". Hurka (2006: 357) says that a theory of value determines "which states of affairs are intrinsically good and which intrinsically evil". Zimmerman (2010) refers to Scanlon (1998: 97) who has called the relationship between valuableness, goodness and intrinsic properties "a buck-passing account, since it "passes the buck" of explaining why something is worthy of being valued from its goodness to some property that underlies it" (Zimmerman 2010: Sec. 2).

[2] A "state of affairs" corresponds to what Chang (1997b), more generally, calls a "bearer of value" or "items".

[3] Sen (2000) suggests a broad conception of states of affairs and refers to "comprehensive outcomes" that include the processes of choice such as actions performed and underlying motivations as well as final outcomes. We define states of affairs more narrowly in order to relate to the established literature.

1979a: 548) or non-preference metrics for individual advantage such as the capability approach (see e.g. Sinnott-Armstrong 2012).

The definition of value only allows to compare states of affairs in terms of their goodness. It does in itself not answer the fundamental question in ethics on how to act. One answer to this question is provided in 'act consequentialism' which says that an action is right if it brings about a state of affairs which is at least as good as each alternative state that results from any other feasible action (Sen 1979b: 464). That is, an action is right if it brings about the best consequences (Hurka 2006: 357). For example, 'act utilitarianism' says that an action is right if it leads to a state of affairs with a larger sum of individual utilities than any other feasible action. That is, act utilitarianism is based on act consequentialism and outcome utilitarianism (Sen 1979b).

Consequentialist ethical theories exclude other ethical theories such as deontological ethics which determines the right action without reference to good consequences or virtue ethics which relies on questions of moral character (Copp 2006: 20). This also means that relationships between values are most relevant for consequentialist ethical theories. For example, if there is a trade-off between values then determining the right action under consequentialism may require balancing of the different degrees of attainment of values with respect to a fundamental value. This raises the further question whether there is a 'fundamental' or intrinsic value. We follow Zimmerman (2010: Sec. 2) to define an intrinsic value as follows:

Definition 2 (Intrinsic value)
An *intrinsic value* is a value which allows a comparison of states of affairs in terms of their intrinsic goodness, i.e. goodness that is not derived from some other goodness.

For example, the intrinsic value in Bentham's utilitarianism (1907 [1789], 1988 [1776]) is aggregate happiness as it is not derived from some other value. A counterexample is an instrumental value that contributes to an intrinsic value.[4,5] For example, Pareto-

[4] On the question whether every value that is not intrinsic is an instrumental value see Zimmerman (2010: Sec. 2)

[5] Chang (1997b: 9) uses a similar framework when she talks about values. For her bearers of value

efficiency is an instrumental value to social welfare as a Pareto-inefficient state can never maximize any function defined over individual utilities. Likewise, political liberties are sometimes seen as instrumental values to ensure other liberties such as civil liberties and liberty of conscience (e.g. in Rawls 1987: 13).

With this philosophical framework one can distinguish between an intrinsic value such as equality in egalitarianism and an instrumental value such as Pareto-efficiency for social welfare. This distinction is important because formulating relationships between instrumental and intrinsic values can be problematic, for instance, if not all intrinsic values are clearly defined. For example, some have questioned whether the trade-off between efficiency and equity is a normatively meaningful relationship because efficiency is an instrumental value and equity is an intrinsic value (Le Grand 1990). In this sense, a trade-off between equity and efficiency begs the question what the intrinsic value behind efficiency is. If the trade-off is assumed to be between intrinsic values, then this requires justification why Pareto-efficiency is an intrinsic value in the same way as equity (Le Grand 1990: 566).

1.2.2 Value pluralism

With regard to intrinsic values, there is the question if there is one single intrinsic value to which all values can be reduced to. The idea that all values can be reduced to one single intrinsic value is called *value monism* (Mason 2011). An example of value monism is Bentham's utilitarianism (1907 [1789], 1988 [1776]) where only aggregate happiness is intrinsically good and all other values are instrumental to this one intrinsic value. Under value monism one can formulate relationships between instrumental values. For example, if social welfare is the intrinsic value this allows trade-offs between different constituents of well-being, such as consumption, leisure or health and education (e.g. in Dasgupta 2005: 241).

(here state of affairs) are compared in terms of their merits (here attributes) with respect to a covering value (here intrinsic value). And a covering value (here intrinsic value) can rely on different contributory values (here instrumental values). She then defines comparativism as the view that justified choice requires the comparison of bearers of values.

If it is impossible to reduce all values to one intrinsic value, this is value pluralism:

Definition 3 (Value Pluralism)
An ethical theory is characterized by *value pluralism* if it is based on multiple intrinsic values.[6]

Value pluralism is found in the work of philosophers such as Isaiah Berlin (Berlin 1969) who speaks of different concepts of liberty and other values such as equality or justice. Also, Brian Barry speaks of a "plurality of 'ultimate' values" (Barry 1965: 5) such as equality and freedom when he discusses political arguments on normative objectives. Ethical theories under value pluralism allow the formulation of relationships between intrinsic values. For example, under value pluralism there can be a trade-off between the intrinsic value of social welfare and the intrinsic value of nature. Similarly, one could think of the relationship between intra- and intergenerational justice in sustainability problems as a relationship between intrinsic values (e.g. Baumgärtner et al. 2012).

Value pluralism can create problems for decision-making such as incomparability which results from incomplete orderings of a set of states of affairs (see Sen 1985: 179, Chang 1997b).[7] For example, it can be that there are two acts that lead to two states of affairs where one has a higher degree of attainment of one intrinsic value such as intergenerational equity, and the other has a higher degree of attainment of another intrinsic value such as intragenerational justice. In order to determine one right action in this context, one cannot use another more 'fundamental' value to balance the degree of attainment of the two values as both are intrinsic values. Nevertheless, there exist

[6] Sen (1985: 178) argues that value pluralism is different from informational pluralism: That is, there can be one intrinsic value (value monism) which subsumes many different kinds of information such as one conception of social welfare that relies on utility and non-preference information. Or there can be multiple intrinsic values (value pluralism) that are all based on the same kind of information such as different social welfare functions that are all based on utility-information.

[7] Hsieh takes 'incommensurability' to refer to the relation between values in the abstract sense and 'incomparability' to refer to the relation between concrete bearers of value which are states of affairs in this analysis (Hsieh 2008: Sec. 1.2).

several approaches to decision-making under value pluralism (e.g. Mason 2011: Sec. 4).[8]

Value pluralism and value monism comprise different ethical theories which has consequences on how the set of feasible states of affairs in terms of values is defined. Ethical theories under value monism view the set of feasible states of affairs in terms of one intrinsic value and multiple instrumental values. This means there can be trade-offs between instrumental values such as equality and economic growth with regard to an intrinsic value such as utilitarian social welfare. Ethical theories under value pluralism view the set of feasible states of affairs in terms of multiple intrinsic value and multiple instrumental values. This means there can be trade-offs between intrinsic values such as social welfare and the intrinsic value of nature. Also, there can be normatively meaningful trade-offs between instrumental and intrinsic values under value pluralism. Yet, these require that all intrinsic values are explicated to allow a clear interpretation of this relationship.

The concrete form of the set of states of affairs plays a big role in the difficulty of decision-making, irrespective of the distinction between value pluralism and value monism. For example, if the feasible set does not force one into making trade-offs between values, decision-making is easy – the higher attainment of one value necessarily increases the attainment of any other value and one can attain a value without compromising any other value. For example, if the feasible set showed that equality and economic growth are always in a win-win relationship, this would also ease political debates on this issue considerably. On the other hand, if the feasible set forces one

[8] Mason (2011: Sec. 4) lists the following approaches: The first approach is practical wisdom (e.g. Anderson 1993, Nagel 1979). Practical wisdom solves problems of incomparability due to multiple intrinsic values without reasoning from general principles, but rather with a faculty of judgement. The second approach uses covering values to determine the weight of each intrinsic value for each circumstance of decision-making (e.g. Chang 1997b). These covering values are said not to be a case of value monism as they always depend on the concrete choice situation. The third approach (suggested by Joseph Raz 1999) who says that one is free to choose among “rationally eligible” options, i.e. states of affairs that are not dominated by others. The fourth approach, says that rather than trying to achieve a rational choice under value pluralism one should accept irresolvable conflicts between values (Williams 1981).

constantly into making trade-offs between values, then decision-making is rather more difficult. In the following, we therefore want to focus our analysis on the feasible set and the relationships between values.

1.3 Formal model and analysis

1.3.1 Actions and states of affairs

In order to relate the model to the discussion on ethics and act consequentialism in particular, we start by defining an action. An *action* $x \in X \subseteq \mathbb{R}^n$ consist of n different projects, so that $x = (x_1, ..., x_i, ..., x_n)$ where x_i denotes the effort spent on project i. Further, superscripts denote different actions, such as x^a and x^b, which differ in terms of the effort spent on projects. The assumption is that projects are given, and that effort is measurable in terms of money, time or some other continuous measure. For example, these projects could be thought of as the projects a government can pursue with a given budget in a given legislative period, e.g. increasing social security payments or increasing investment in infrastructure.[9]

A *state of affairs* $y \in Y \subseteq \mathbb{R}^m$ is a complete description of the world in terms of m different attributes, so that $y = (y_1, ..., y_j, ..., y_m)$. Thus, the set Y can be thought of as the set of feasible states of affairs in terms of attributes. These attributes y_j contain continuously measurable information on, for example, stocks such as physical, social, human and natural capital; on flows such as payoffs, income, and environmental (dis-)services; and on individuals such as their number, the distribution of income, capabilities, and resources over individuals. We assume that the set of states of affairs Y is non-empty, closed and bounded and, thus, a compact set. Further, superscripts denote different states of affairs, such as y^a and y^b. Here, the assumption is that all these attributes are continuously measurable or at least representable, prima facie, via indicators, which allow to measure distance in Y-space.[10] For example, a state of affairs

[9] This would imply the familiar budget constraint in terms of money $X = \{x | \sum_i x_i \leq m\}$

[10] This excludes the use of discrete attributes as in e.g. Pattanaik and Xu (2012).

could contain attributes on the income distribution or the level of gross domestic product (GDP).

The *outcome function* $F : X \rightarrow Y$ with $y = F(x)$ shows what state of affairs y results from an action x, where the action-space X is the domain and the state-of-affairs space Y is the range of the function. Furthermore, $F(x)$ is deterministic and assumes technical efficiency, in that an action results in the best possible change in an attribute.[11] For example, the outcome function could show how social security payments affect the income distribution and how investment in infrastructure affects GDP.

1.3.2 Value relation and single-value index function

As discussed in Section 1.2 different ethical theories determine values differently. Therefore, the *value set* V contains $l = \#V$ values where $v \in V$ denotes a value in the set V. The concrete nature of these values and the number l of values are determined within a given ethical theory. In this way, the value set should only contain values which allow meaningful relationships (see Section 1.2). For example, under value monism a value set V with $l = 1$ contains one intrinsic value such as utilitarian social welfare. In the same way, V could contain multiple values $l > 1$ under value monism if there are different instrumental values to the single intrinsic value. Alternatively, under value pluralism, V with $l > 1$ could contain different intrinsic values. As we are interested in relationships between values, formulations of value sets with more than one value are most relevant to our analysis.

For the comparison of states of affairs there exists a value relation for each value.

Definition 4 (Value relation)
A *value relation* $\succeq_v$ is a binary relation on the set of states of affairs Y with respect to a value $v \in V$, where $y^a \succeq_v y^b$ means that y^a is as at least as good as y^b in terms of value v.

[11] If the outcome function is invertible this allows the identification of a specific action from a given state of affairs, which allows the consequentialist evaluation of actions. By assuming the invertibility of the outcome function, we can focus on states of affairs in most of the subsequent analysis.

For example, under egalitarianism income inequality could provide a value relation for the value of equality. In economics the properties of value relations are discussed in the social choice and fair social orderings literature (e.g. Fleurbaey et al. 2005) Also, in philosophy there has been some discussion on the properties of value relations in recent years (e.g. Rabinowicz 2008). The traditional framework includes three value relations: *better than*, *worse than* and *equal to* ("trichotomy thesis") (Chang 1997b: 4).[12] These relations can be represented by the value relation given above as follows.

Definition 5 (Trichotomy in value relations)
The following relations can be derived from $\succeq_v$:

- *better than*: $y^a \succ_v y^b$ if and only if $y^a \succeq_v y^b$ and not $y^b \succeq_v y^a$
- *worse than*: $y^a \prec_v y^b$ if and only if $y^b \succeq_v y^a$ and not $y^a \succeq_v y^b$
- *equal to*: $y^a \sim_v y^b$ if and only if $y^a \succeq_v y^b$ and $y^b \succeq_v y^a$

The properties of value relations are directly relevant for decision-making under act consequentialism as these determine the ordering of states of affairs provided by a value relation (see e.g. Sen 1970b: 9). Common assumptions on the properties of value relations are the following:

Definition 6
A value relation $\succeq_v$ is

- *reflexive*, if and only if for all $y^a \in Y$ one has $y^a \succeq_v y^a$
- *transitive*, if and only if for all $y^a, y^b, y^c \in Y$ if $y^a \succeq_v y^b$ and $y^b \succeq_v y^c$, then $y^a \succeq_v y^c$
- *complete*, if and only if for all $y^a, y^b \in Y$ one has $y^a \succeq_v y^b$ or $y^b \succeq_v y^a$ (or both)

[12] Chang (1997a, 2002a, 2005) has suggested a fourth value relation that she called "on a par" where two states of affairs are comparable, yet not better than, worse than or equal to another. Boot (2009) comments that this debate is not directly relevant to justified choice as it creates the same problems as incomparability between bearers of value where no value relation holds. Thus, for the course of this analysis we will not consider parity as a fourth value relation.

- *anti-symmetric*: if and only if for all $y^a, y^b \in Y$ one has $y^a \succeq_v y^b$ and $y^b \succeq_v y^a$ yields $y^a = y^b$
- *continuous*, if and only if for any sequence of pairs $\{(y_n^a, y_n^b)\}_{n=1}^{\infty}$ with $y_n^a \succeq_v y_n^b$ for all n, $y^a = \lim_{n\to\infty} y_n^a$, and $y^b = \lim_{n\to\infty} y_n^b$, one has $y^a \succeq y^b$

Each of these assumptions on value relations can be, and has been, criticized in philosophy. For example, Broome (1991: 18) considers reflexivity and transitivity as the prime characteristics of rationality. Hausman (1993) criticizes just these characteristics with respect to value relations due to the vagueness of some normative predicates. The assumption of completeness excludes cases where no value relation holds between states of affairs and, therefore, excludes problems of incomparability and incommensurability which are sometimes highlighted in philosophical reasoning (e.g. Chang 1997a, Hsieh 2008). Still, completeness of each respective value relation allows us to model incomparability. That is, two states of affairs $y^a, y^b \in Y$ can be incomparable in terms of values, if two intrinsic values $v, w \in V$ individually can provide a complete ordering of states of affairs, but differ in their ordering regarding of the two states, e.g. $y^a \succ_v y^b$ and $y^a \prec_w y^b$. Then, both states of affairs are ordered completely with respect to $v, w \in V$, but y^a is better than y^b in terms of v and worse in terms of w, so that the two states of affairs are incomparable in the sense that it is impossible to rank one over the other.

A prominent discussion on a complete and transitive value relation is the social ordering in Arrow's impossibility theorem (Arrow 1951). It shows the impossibility of establishing a complete and transitive social choice function from ordinal non-comparable utility information and further axioms (see e.g. Roemer 1996). Still, there are different interpretations of Arrow's impossibility theorem. A rather optimistic one is provided by Sen (1999) who highlights the different possibilities of making interpersonal comparisons in terms of incomes, primary goods, or basic needs. The difficulty in finding and constructing such complete and transitive social orderings for the social evaluation of states of affairs has long been debated for example in the related discussion on indices for distributive justice (e.g. Fleurbaey 2007). A salient example is Rawls's (1971) theory of justice which uses primary goods as its metric of individual advantage to identify the worst-off individuals which creates the need to construct a suitable ordering for the comparison of individuals.

The more technical assumption of continuity excludes some value relations that might be relevant in ethical theories. For example, lexicographic value relations are not continuous as they give priority to an increase in one attribute irrespective of any increases in some another attribute.

Definition 7

A value relation $\succeq_v$ is *lexicographic* for two attributes[13], if there exist two attributes $y_i, y_j \in \{y_1, ..., y_m\}$ for which $y^a \succeq_v y^b$ if and only if

$$\begin{gathered} y_i^a > y_i^b \\ \text{or } y_i^a = y_i^b \text{ and } y_j^a \geq y_j^b. \end{gathered}$$

An example for this is John Ralws lexicographic ordering of basic liberties before prosperity in his difference principle (Rawls 1971: Sec. 8) or James Griffin's (1986: 83) discussion of cases where one attribute trumps another. Furthermore, this lexicographic value relation is reflexive, transitive and complete.

In the following, we make the following assumptions on value relations:

Assumption 1

The value relation $\succeq_v$ is reflexive, transitive and complete for all $v \in V$.

These assumptions establish an ordering (as in Sen 1970b: 9) for each value $v \in V$. These orderings allow a maximizing account of consequentialism which says an action is right if it leads to a state of affairs with the highest possible degree of attainment of a value (e.g. Broome 1991, Sinnott-Armstrong 2012). The assumption of monotonicity, that more of an attribute is always better, is not required here. For example, if there is satiation with respect to some attributes for a value relation, this value relation can still be reflexive, transitive and complete.

[13] This is a very special lexicographic ordering of attributes but which still captures the intuition on priority of an increase in one attribute over any increase in other attributes. A more general treatment of lexicographic relations can be found in e.g. Hougaard and Tvede (2001).

Under certain further assumptions on value relations there exists an index function that represents the ordering of a value relation.

Definition 8 (Index function)
A function $I_v : Y \to \mathbb{R}$ is an *index function* representing value relation $\succeq_v$ if and only if, for any $y^a, y^b \in Y$,

$$y^a \succeq_v y^b \Leftrightarrow I_v(y^a) \geq I_v(y^b)$$

The index function I_v is an index of value v and shows the degree of attainment of this value.

Proposition 1 (Existence of index function)
If the value relation $\succeq_v$ is complete, transitive and continuous on Y, then there exists a continuous index function I_v: $Y \to R$ which represents the value relation $\succeq_v$

Proof. By analogy the standard proof for the existence of a utility function applies, see e.g. MasColell et al. (1995: 47). □

An example of such an index function is a utility function which show the degree of attainment of the value of an individual preference. Another example are Bergson-Samuelson social welfare functions $W(\cdot)$ which show the degree of attainment of social welfare as a function of individual utilities U_i (e.g. Suzumura 1987). Despite Arrow's impossibility theorem and the associated debate on its axiomatic foundation, social welfare functions that are based on interpersonally comparable utility information are used in areas of applied economics such as the economics of climate change (e.g. Stern 2007).

1.3.3 Value pluralism and value-efficiency

As discussed in Sections 1.2 and 1.3.1, the value set V can contain one single or multiple values depending on the respective ethical theory. For example, it can contain one intrinsic value under value monism or multiple intrinsic values under value pluralism. Given such a value set V one can define value-efficiency as follows:

Definition 9 (Value-efficiency)
A state of affairs $y \in Y$ is *value-efficient* if and only if there exists no other state of affairs $y' \in Y$ for which $y' \succeq_v y$ for all $v \in V$ and $y' \succ_v y$ for at least one $v \in V$.

This criterion is in line with the maximizing account of consequentialism, as discussed above, where a state of affairs with a higher attainment of one value is always better than one with a lower degree of attainment. Indeed, value-efficiency provides some guidance for actions as it characterizes those actions as bad that lead to value-inefficient states of affairs. As value-efficiency itself is not a very demanding concept it provides room for many consequentialist ethical theories that are based on different definitions of value. Some authors have gone further and suggested value-efficiency as a criterion for choosing states of affairs under value pluralism. For example, Raz (1997) says that under intrinsic value pluralism any value-efficient state of affairs can be chosen with free will (what he calls 'rationally eligible' states of affairs). Similarly, Brun and Hirsch-Hadorn (2008) determine value-efficient states of affairs concerning the multiple values within the sustainability concept.

In economics this notion of value-efficiency has also been mentioned and discussed. For example, Sen (1979a: 553) refers to 'dominance' with respect to normative considerations and discusses its difference to Pareto-efficiency under value pluralism. Le Grand (1990: 559) specifically uses the notion of value-efficiency in his discussion of the equity-efficiency trade-off. Pattanaik and Xu (2012) discuss dominance with respect to evaluative attributes and the use of context-specific information in decision-making. Others have noted the limited usefulness of value-efficiency for decision-making as it provides only incomplete orderings of states of affairs (Sen 1985: 178). Finally, Dasgupta (2005) shows the range of values that can be incorporated in economic analysis by distinguishing Pareto-efficiency which is based solely on utility information from broader 'efficiency' which is based on utility and non-preference information. Based on these specific contributions, we will further discuss the relationship between Pareto-efficiency and value-efficiency in Section 1.4.

With Definition 9 of value-efficiency, one may ask under what conditions there exist value-efficient states of affairs for different values.

Proposition 2 (Existence of value-efficient state of affairs)
There exists at least one value-efficient state of affairs if the set of states of affairs Y is a finite set and $\succeq_v$ satisfies Assumption 1 for all $v \in V$, or if the set of states of affairs Y is an infinite set and every $\succeq_v$ with $v \in V$ satisfies either (i) or (ii):

(i) $\succeq_v$ satisfies Assumption 1 and continuity.[14]

(ii) $\succeq_v$ is lexicographic.[15]

Proof. See Appendix A.1 □

This shows that there exists a value-efficient state of affairs for many different combinations of values and sets of states of affairs. For example, there exists a value-efficient state if some values are lexicographic, such as priority of the worst-off, and others are continuous, such as income inequality.

1.3.4 Relationships between values

As argued in Section 1.2 the form of the set of feasible states of affairs and relationships between values impact on the difficulty of decision-making. This leads us to distinguish the following relationships[16] between two values. Also, for value sets with more than two values $l > 2$, we will continue with pairwise relationships between two values.

Definition 10 (Relationships between values)
In a feasible state of affairs y a *relationship* R between two values $v, w \in V$ is a binary

[14] There exists a broad literature on the existence of maximal elements with weaker continuity and transitivity axioms following Bergstrom (1975) and Walker (1977) As we use index functions in parts of the subsequent analysis, (i) refers to the standard continuity axiom.

[15] The existence of maximal elements for lexicographic relations is treated more generally in e.g. Hougaard and Tvede (2001) and Houy and Tadenuma (2009)

[16] These relationships are common in many discussions, e.g. Le Grand (1990), Engel et al. (2008), Glotzbach and Baumgärtner (2012), and Baumgärtner et al. (2012).

relation on $V \times V$, denoted as vRw, with the following special cases:[17]

(i) R is a *trade-off* relationship if and only if $y' \succ_v y$ implies $y' \prec_w y$ for all $y' \in Y$

(ii) R is a *win-win* relationship if and only if $y' \succ_v y$ implies $y' \succ_w y$ for all $y' \in Y$

(iii) R is an *independence* relationship if and only if there exists $y', y'' \in Y$ such that $y' \succ_v y$ implies $y' \succ_w y$ and $y'' \succ_v y$ implies $y'' \sim_w y$

The relationships R are defined as directed relationships from one value v on another value w. The converse relationship only holds if R is symmetric.[18] Definition 10 does not include incomparability as a relationship, which is due to the completeness of value relations in Assumption 1.

The most familiar relationship (i) is the one of a *trade-off* between values. A trade-off is the case when attaining one value to a higher degree necessarily reduces the degree to which one attains the other one. This is a symmetric relationship which holds both ways. An example from the introduction is the trade-off between efficiency and equality.

The next relationship (ii) is the one of a *win-win relationship* between values. A win-win relationship is the case where achieving one value facilitates achieving the other one, that is, attaining one value to a higher degree induces a higher degree of attainment of the other one. This is an asymmetric relationship which does not necessarily hold both ways. An example from the introduction is the win-win relationship between income equality and intergenerational income mobility.

The last relationship (iii) is one of *independence* between values. Independence is the case when values can be achieved independently, that is, attaining one value to a higher degree does not necessitate any change in the degree to which one attains the other one. An example is the case where the intragenerational distribution of CO_2 emission permits

[17] More generally, relationships may be defined for a local environment around the state of affairs $y \in Y$ in which the condition(s) must hold. This leads to the additional condition on y' that $\|y' - y\| \leq \varepsilon$ for $\varepsilon > 0$.

[18] A trade-off relationship is always symmetric in that one value trades-off with another. Win-win relationships and independence relationships are not symmetric. For example, it can be that relationships are only onesided as illustrated below in Figure 1.2.

is independent from the intergenerational distribution of CO_2 emission permits, so that the attainment of intragenerational distributive justice is independent of the attainment of intergenerational distributive justice.

These relationships are illustrated in Figure 1.2 for the case of a feasible set of states of affairs (depicted in value-space) with two values $v, w \in V$.[19] The axes in this figure show the degree of attainment of values v and w via the respective index functions I_v and I_w. A point in this figure represents a state of affairs such as y^b which results from a specific action such as x^b. The black boundary line is the frontier which delimits states of affairs that are feasible, those inside the grey area, from states of affairs that are not feasible, those inside the white area. The north-eastern part of the frontier, between y^c and y^d shows all value-efficient states of affairs. In y^a there is an independence relationship between values v and w as the attainment of either value can be increased *without necessarily changing* the attainment of the other. In y^e there is a win-win relationship between values v and w as an increased attainment of w *necessarily increases* the attainment of v. Yet, this win-win relationship is only onesided as the attainment of v can be increased without necessarily increasing the attainment of w. In y^f the opposite case of win-win relationship is depicted. In any state of affairs along the frontier between y^c and y^d there is a trade-off relationship between the attainment of values v and w as attaining either one to a higher degree *necessarily reduces* the degree of attainment of the other one.

Figure 1.3 illustrates the relationships between values for the case of a feasible set of states of affairs (again depicted in value-space) with three values $u, v, w \in V$. Correspondingly, there are three axes in this figure which show the degree of attainment of values u, v and w via the respective index functions I_u, I_v and I_w. The surface of the set in value-space, which is indicated by the thin black lines, delimits states of affairs that are feasible, those inside the grey volume, from states of affairs that are not feasible, those inside the white volume. Also, the surface shows all value-efficient states of

[19] The shape of the set of feasible states of affairs Y in value-space as in Figures 1.2 and 1.3 relies on further assumptions regarding the feasible set Y in attribute-space that go beyond the ones made in Section 1.3.

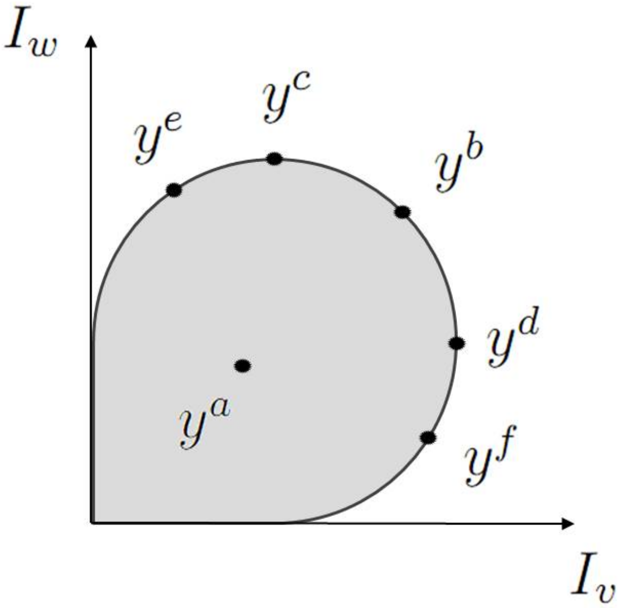

Figure 1.2: Convex set of feasible states of affairs in value-space for two values $v, w \in V$. The axes show the degree of attainment of values v and w via the respective index functions I_v and I_w. Each point in the figure corresponds to a state of affairs that results from a specific action. The black frontier delimits feasible states of affairs in the grey area from not feasible ones in the white area. Value-efficient states of affairs are on the frontier between y^c and y^d. There are different relationships in different states of affairs: Trade-off relationship in y^c,y^b,y^d; Win-win relationship in y^e,y^f; Independence relationship in y^a. Adapted from Baumgärtner et al. (2012).

affairs. In y^a there is a trade-off relationship between v and w as increasing either one *necessarily reduces* the attainment of the other value. Also, there is an independence relationship between v and u as increasing the degree of attainment of v can either lead to an increased attainment of u (by going from y^a to y^c) or remain on the same level of attainment (by going from y^a to y^b). There are no win-win relationships in Figure 1.3 as in no state of affairs the increased attainment of one value *necessarily increases* the attainment of any other.

From Figures 1.2 and 1.3 the connection between a trade-off between values and

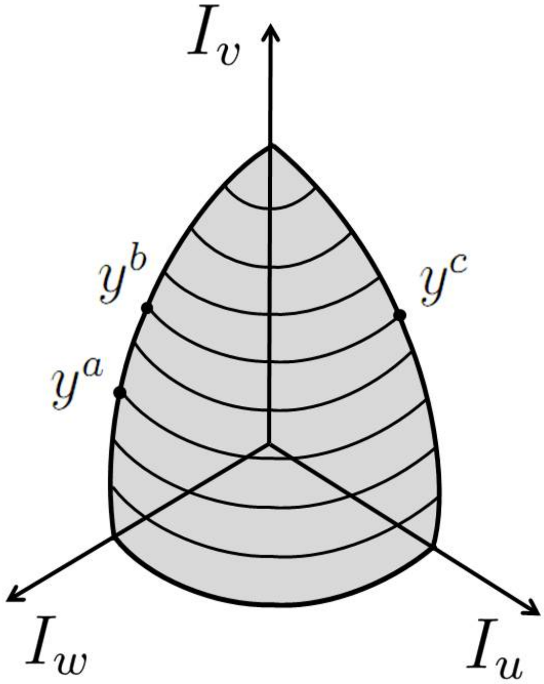

Figure 1.3: Convex set of feasible states of affairs in value-space for three values $u, v, w \in V$. The three axes show the degree of attainment of values u, v and w via the respective index functions I_u, I_v and I_w. Each point in the figure corresponds to a state of affairs that results from a specific action. The surface of the set in value-space, indicated by the thin black lines, delimits states of affairs that are feasible in the grey volume from states of affairs that are not feasible in the white volume. The surface also contains all value-efficient states of affairs. There are different relationships in between different values in a state of affairs: In y^a there is a trade-off relationship between v and w and independence relationship between u and v.

efficiency becomes apparent:

Proposition 3 (Value-efficiency and relationships between values) (i) If there exist at least two value-efficient state of affairs, then in every value-efficient state of affairs there exists at least one trade-off relationship between two values $v, w \in V$.

(ii) If in a state of affairs $y \in Y$ there is a trade-off between at least two values $v, w \in V$, then y is value-efficient.[20]

(iii) If there exist at least three values in the value set, $\#V = l \geq 3$, there can be win-win relationships or independence relationships between values in a value-efficient state of affairs.

Proof. See Appendix A.2 □

This means that win-win relationships between two values do not necessarily indicate value-inefficient states of affairs. Rather, the absence of value trade-offs is a clear indication of a value-inefficient state of affairs. For example, suppose there are three values such as income equality, intergenerational income mobility and overall economic well-being as indexed by GDP. There can be a win-win relationship between income equality and intergenerational income mobility in a value-efficient state of affairs. Yet, it must be that there is a trade-off with GDP and economic well-being as this state of affairs would otherwise not be value-efficient. If the value set contains only two values $\#V = l = 2$ then there is always a trade-off between these values in any value-efficient state of affairs. This can be illustrated in Figure 1.2 where the increased attainment of one value reduces the attainment of the other in all value-efficient states on the frontier between y^c and y^d.

1.4 Value-efficiency and Pareto-efficiency

1.4.1 Individualistic framework

As the definition of value-efficiency and relationships between values is based on a very broad consequentialist definition of value, it is interesting to look at its connection to the familiar definition of Pareto-efficiency. This requires to introduce the familiar individualistic framework from welfare economics. In a first step, define a preference

[20] If one considers locally defined relationships this may not hold as it may be that in a value-inefficient state there are trade-offs in a local environment.

relation as special case of a value relation with respect to the value of an individual preference. Suppose there are z individuals whose self-regarding preferences p_k (with $k \in \{1, ..., z\}$) are contained in a set P with $\{p_1, ..., p_k, ..., p_z\} = P$.

Definition 11 (Preference relation)
For all individuals $k = 1, ..., z$, a *preference relation* $\succeq_{p_k}$ of individual k is a binary relation on the set of states of affairs Y, where $y^a \succeq_{p_k} y^b$ means that y^a is as at least as good as y^b in terms of individual preference p_k.[21]

In this setting, Pareto-efficiency can be defined as follows:

Definition 12 (Pareto-efficiency)
A state of affairs $y \in Y$ is *Pareto-efficient* if and only if there exists no other state of affairs $y' \in Y$ for which $y' \succeq_{p_k} y$ for all individual preferences $p_k \in P$ and $y' \succ_{p_k} y$ for at least one individual preference $p_k \in P$.

This definition of efficiency shows the liberal heritage of Pareto-efficiency which is based on the set P of individual preferences and the idea that all individuals should be free to pursue the satisfaction of their individual preferences.

1.4.2 Value sets and individual preferences

In the next step, we consider four different cases for the role of preferences in value sets to discuss the connection between value-efficiency and Pareto-efficiency. The first case is the one where the value set contains all individual preferences and only individual preferences $P = V$ in which case value-efficiency in Definition 9 reduces to Pareto-efficiency in Definition 12.

[21] The assumption that individuals have complete preferences over the whole set of state of affairs Y where each state of affairs $y \in Y$ has m different attributes is stronger than the usual assumption that individuals only have preferences over their individual consumption sets. While it is possible to define preference relations over only particular attributes of a state of affairs that concern an individuals self-regarding preference, we do not do so here in order to keep formal requirements to a reasonable level that sill allows us to support our results.

Proposition 4 (Equivalence of value-efficiency and Pareto-efficiency)
Value-efficiency is equivalent to Pareto-efficiency if and only if $P = V$.

Proof. See Appendix A.3 □

The equivalence of value-efficiency and Pareto-efficiency is briefly discussed, for example, in Dasgupta (2005: 240) who discusses the case when all non-preference values are incorporated into a notion of individual utility. This case is well studied in welfare economics where Pareto-efficiency serves as a central normative criterion, especially following Arrow's impossibility theorem (Sen 1999: 352). Several authors have defended this prominent use of Pareto-efficiency. For example, Buchanan says that "[s]ince 'social' values do not exist apart from individual values in a free society, consensus or unanimity (mutuality of gain) is the only test which can insure that a change is beneficial" (Buchanan 1959: 137). This exclusive reliance on individual preferences in the assessment of states of affairs leads to the debate on the ethical appeal of welfarism (e.g. Sen 1979b, Sen 1980, Roemer 1996, Kaplow and Shavell 2001, Fleurbaey et al. 2003). For example, this concerns the question if all values can be reduced to individual preferences and how values that deviate from individual preferences can be justified. Further, if the value set contains only individual preferences, the earlier results in Section 1.3.4 can be interpreted accordingly. For example, Proposition 3 says that there can be a win-win relationships between two individual preferences in Pareto-efficient states of affairs when there are trade-offs between other individual preferences.

The second case where only some individual preferences are included in the value set, $V \subset P$, is discriminatory. This would require an ethical theory where preferences of some people do not matter in the assessment of states of affairs. This of course includes cases where the value set does not contain all individual preferences and further non-preference values.

There may be values that are not derived from individual preferences, sometimes called non-preference (or non-utility) values.

Definition 13 (Non-preference value)
A value v is a non-preference value if $v \in V \backslash P$.

For example, Dasgupta (2005: 240) mentions democracy and civil liberties as examples of such additional non-preference values. Another salient example, he mentions, is the liberal paradox from microeconomic theory where under certain conditions liberalism (as an inviolable personal domain) is in conflict with Pareto-efficiency (Sen 1970a).

Non-preference values allow the third case of value sets which are based ethical theories, especially the ones labeled 'non-welfarist', which do not base their value set V on the satisfaction of individual preferences, $P \cap V = \varnothing$. For example, Kaplow (2007) discusses the capability approach by Sen (1980) and primary goods in Rawls (1971) as theories do not include individual preferences but rather the means to the fulfillment of individual preferences in their assessment of states of affairs. He goes on to discuss the conflict of these alternative ethical theories with Pareto-efficiency. Similarly, there is a discussion on the conflict of efficiency criteria for value sets derived from different ethical theories (e.g. Brun and Tungodden 2004, Fleurbaey 2007). This concerns the conflict between an efficiency criterion for one value set V (value-efficiency) which contains only non-preference values such as primary goods and another efficiency criterion on a set P (Pareto-efficiency) which contains only individual preferences.

A related case is the fourth case where the value set contains individual preferences and additionally non-preference values, $P \subset V$. In this case Pareto-efficiency and value-efficiency are not equivalent. Still, the inclusion of non-preference values into an economic framework can lead to a conflict with Pareto-efficiency in some sets of feasible states of affairs (e.g. Kaplow and Shavell 2001, Fleurbaey et al. 2003).

This last case can be illustrated in Figure 1.4. It shows a convex set of feasible states of affairs in value-space with three values: two individual preferences $p_1, p_2 \in P \subset V$ and one non-preference value $v \in V$. The axes in this figure show the degree of attainment of the three values via their respective index functions: two utility functions U_1, U_2 and an index function for a non-preference value I_v. The surface of the set in value-space, which is indicated by the thin black lines, delimits states of affairs that are feasible, those inside the grey volume, from states of affairs that are not feasible, those inside the white volume. Also, the surface shows all value-efficient states of affairs. All Pareto-efficient states of affairs are those on the dashed line on the surface where the non-preference value is attained to the lowest degree. Thus, surface of the set in value-space encompasses all value-efficient and all Pareto-efficient states of affairs.

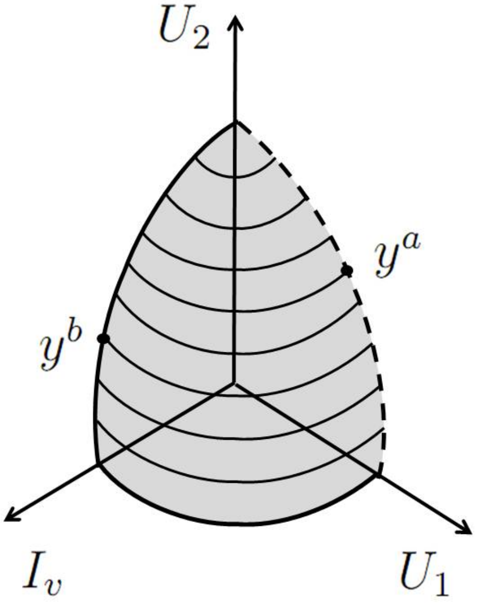

Figure 1.4: Convex set of feasible states of affairs in value-space with non-preference values and preferences. The axes in this figure show the degree of attainment of the three values via their respective index functions: two utility functions U_1, U_2 and an index function for a non-preference value I_v. The surface of the set in value-space, which is indicated by the thin black lines, delimits states of affairs that are feasible, those inside the grey volume, from states of affairs that are not feasible, those inside the white volume. All value-efficient states of affairs are on the surface of the set in value-space. All Pareto-efficient states of affairs are those on the dashed line on the surface where the non-preference value is attained to the lowest degree. State of affairs y^a is Pareto-efficient and value-efficient, state of affairs y^b is value-efficient and not Pareto-efficient.

Here, one can distinguish a Pareto-efficient and value-efficient state of affairs such as y^a from a value-efficient state such as y^b. States of affairs that are not on the surface are value-inefficient and Pareto-inefficient. In the following we focus on this case of preferences and non-preference values $P \subset V$ as this has attracted the most interest in economics.

Corollary 1 (to Proposition 3 (iii))
If there are at least two non-preference values $v, w \in V \backslash P$ and at least two individual preferences $p_1, p_2 \in P \subset V$, then there can be any relationship between the two values v, w in a Pareto-efficient state of affairs $y \in Y$.

Proof. See Appendix A.4 □

This result is relevant to the interpretation of Proposition 3 as this means that there can be a win-win relationship between non-preference values in a given Pareto-efficient state of affairs. For example, if individuals do not care about the values of equality and liberty as a minimal inviolable personal domain (as in Sen 1970a), then there can be a Pareto-efficient state which exhibits a win-win relationship between these non-preference values.

Proposition 5 (Conflict of value-efficiency and Pareto-efficiency)
If $P \subset V$, then there may exist a value-efficient state of affairs that is not Pareto-efficient, and vice versa, a Pareto-efficient state that is not value-efficient.[22]

Proof. See Appendix A.5 □

Generally, a Pareto-efficient state of affairs is not value-efficient if there exists another state of affairs where people are equally well off and some non-preference value is attained to a higher degree. For example, Dasgupta (Dasgupta 2005: 240) considers the case of the liberal paradox, where a state that respects liberty can be Pareto-inefficient, but is value-efficient as illustrated in Figure 1.4 through state of affairs y^b.

[22] This is a comparatively weak statement. Kaplow and Shavell (2001) make a stronger statement yet they refer to the set of all conceivable states, and not explicitly to the set of feasible states. It is clear that a general statement regarding the conflict between value and Pareto-efficiency in the feasible set requires much more detailed assumptions regarding values and the feasible set as the broad literature on in social choice theory shows.

1.5 Conclusion

We studied relationships between values on the set of feasible states of affairs by following a consequentialist definition of value that focuses on the assessment of states of affairs. We showed that for value sets with three or more values, there can be independence, win-win relationships along with trade-offs between values in value-efficient states of affairs. That is, neither win-win nor independence relationships between values indicate value-inefficient states of affairs, if there are more than two values. We saw that in any case the absence of trade-offs between values indicates a value-inefficient state of affairs. For ethical theories that determine values that are not reducible to individual preferences, we saw that there can be win-win relationship between non-preference values in a Pareto-efficient state of affairs.

Regarding the connection between Pareto-efficiency and value-efficiency, we showed that the former is a special case of the latter when individual preferences are taken to be the only values. This underscores how important the discussion on conceptions of well-being is for the analysis of conflicts of non-preference values with Pareto-efficiency as these conceptions determine what values are incorporated in individual preferences and which are not (e.g. Sen 1980, Anderson 1993).

While these results are quite general, they are not derived for a concrete ethical problem that generates a specific set of feasible states of affairs. That is, in a detailed economic model results on relationships between particular values could be derived from more specific assumptions on individual preferences, instruments and resources. In this vein, the general insights into value-efficiency and relationships between values in this paper could be applied for concrete ethical theories such as Sen's capability approach (e.g. Pattanaik and Xu 2012) or the relationship between intra- and intergenerational justice (e.g. Baumgärtner et al. 2012).

More generally, the analysis of pairwise relationships between two values proved limited for cases where more than two values are involved. For example, in this case the information on win-win or trade-off relationships between two values must be interpreted much more carefully in light of the effects other potentially involved values. A possibility would be to define binary relationships between values with the provision that *the degree of attainment of all other values remains constant*. This would ensure that there are only trade-offs between values in a value-efficient state of affairs.

Acknowledgements

We are grateful to Dave Abson, Urs Allenspacher, Georg Brun, John-Oliver Engler, Gertrude Hirsch-Hadorn, Ivo Wallimann-Helmer, Michael Liebendörfer, Dominic Roser, Michael Schefczyk, Fabian Schuppert, Martin Quaas, the team of the Chair of Economics / Resource Economics at ETH Zürich, for critical and constructive discussion, and to the German Federal Ministry of Education and Research for financial support under grant 01UN1011A. Nikolai Hoberg would like to thank for discussion and the hospitality at the Chair of Economics / Resource Economics at ETH Zürich from March to June 2012.

Bibliography

Alesina, A., Di Tella, R., and MacCulloch, R. (2004). Inequality and happiness: are Europeans and Americans different? *Journal of Public Economics*, 88:2009–2042.

Anderson, E. (1993). *Value in Ethics and Economics.* Harvard University Press, Cambridge, Mass.

Arrow, K. (1951). *Social Choice and Individual Values.* Wiley, New York.

Barry, B. (1965). *Political Argument.* Routledge and Keagan Paul, London.

Baland, J.M, Bardhan, P., Das, S., Mookherjee, D., and Sarkar, R. (2010). The environmental impact of poverty: Evidence from firewood collection in rural nepal. *Economic Development and Cultural Change*, 59(1):23–61.

Baumgärtner, S., Glotzbach, S., Hoberg, N., Quaas, M.F., Stumpf, K. (2012). Economic analysis of trade-offs between justices. *Intergenerational Justice Review*, 1:4–9.

Bergstrom, T. (1975). Maximal elements of acyclic relations on compact sets. *Journal of Economic Theory*, 10(3):403–404.

Bentham, J. (1907). *An Introduction to the Principles of Morals and Legislation.* Clarendon Press, Oxford.

Bentham, J. (1988). *A Fragment on Government.* Cambridge University Press, Cambridge.

Berlin, I. (1969). *Four Essays on Liberty.* Oxford University Press, Oxford.

Björklund, A., and Jäntti, M. (1997). Intergenerational income mobility in Sweden compared to the United States. *American Economic Review*, 87(5):1009–1018.

Blank, M.R. (2002). Can equity and efficiency complement eachother? *Labour Economics*, 9(4):451–468.

Boot, M. (2009). Parity, incomparability and rationally justified choice. *Philosophical Studies*, 146(1):75–92.

Broome, J. (1991). *Weighing Goods. Equality, Uncertainty and Time.* Basil Blackwell Inc., Cambridge, Mass.

Brun, B. and Tungodden, B. (2004). Non-welfaristic theories of justice: Is "the intersection approach" a solution to the indexing impasse? *Social Choice and Welfare*, 22(1):49–60.

Brun, G. and Hirsch-Hadorn, G. (2008). Ranking policy options for sustainable development. *Poiesis and Praxis*, 5(1):15–31.

Buchanan, J.M. (1959). Positive economics, welfare economics, and political economy. *Journal of Law and Economics*, 2:124–138.

Chang, R.(ed.) (1997a). *Incommensurability, Incomparability and Practical Reason.* Harvard University Press, Cambridge, Mass.

Chang, R. (1997b). Introduction. In: Chang, R. (ed.), *Incommensurability, Incomparability and Practical Reason*, Harvard University Press, Cambridge, Mass. 1–34.

Chan, R. (2002 a). The Possibility of Parity. *Ethics*, 112(4):659–688.

Chan, R. (2002 b). *Making Comparisons Count.* Routledge, New York.

Chang, R. (2005). Parity, interval value and choice. *Ethics*, 115(2):331–350.

Copp, D. (2006). Introduction: Metaethics and normative ethics. In: Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, Oxford. 357–359.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102.

Cowen, T. (2007). The importance of defining the feasible set. *Economics and Philosophy*, 23(1):1–14.

Dasgupta, P. (2005). What do economists analyze and why: values or facts? *Economics and Philosophy*, 21(2):221–278.

De Haan, J., Lundström, S., and Sturm, J.-E. (2006). Market-oriented institutions and policies and economic growth: a critical survey. *Journal of Economic Surveys*, 20(2):157–191.

Engel, S., Pagiola, S., and Wunder, S. (2008). Designing payments for environmental services in theory and practice: An overview of the issues. *Ecological Economics*, 65(4):663–674.

Fleurbaey, M., Tungodden, B., and Chang, H. (2003). Any non-welfarist method of policy assessment violates the Pareto principle: A comment. *Journal of Political Economy*, 111(6):613–639.

Fleurbaey, M., Suzumura, K., and Tadenuma, K. (2005). The informational basis of the theory of fair allocation. *Social Choice and Welfare*, 24(2):311–341.

Fleurbaey, M. (2007). Social choice and the indexing dilemma. *Social Choice and Welfare*, 29(4):633–648.

Glotzbach, S., and Baumgärtner, S. (2012). The relationship between intragenerational and intergenerational ecological justice. *Environmental Values*, 21(3):331–335.

Griffin, J. (1986). *Well-Being: Its Meaning, Measurement and Moral Importance.* Clarendon Press, Oxford.

Hausman, D. (1993). Structure of good. A review of weighing goods by John Broome. *Ethics*, 103(4):792–806.

Hausman, D. and McPherson, M. (1993). Taking ethics seriously: Economics and contemporary moral philosophy. *Journal of Economic Literature*, 31(2):671–731.

Hougaard, J. and Tvede, M. (2001). The existence of maximal elements: generalized lexicographic relations. *Journal of Mathematical Economics*, 36(2):111–115.

Houy, M. and Tadenuma, K. (2009). Lexicographic compositions of multiple criteria for decision making. *Journal of Economic Theory*, 144(4):1770–1782.

Hsieh, N. (2008). Incommensurable values. In: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2008 Edition, [Online available at: `http://plato.stanford.edu/archives/fall2008/entries/value-incommensurable/`]

Hurka, T. (2006). Value Theory. In: Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, Oxford. 357–359.

Kaplow, L. (2007). Primary goods, capabilities,... or well-being? *The Philosophical Review*, 116(4):603–632.

Kaplow, L. and Shavell, S. (2001). Any non-welfarist method of policy assessment violates the Pareto principle. *Journal of Political Economy*, 109(2):281–286.

Le Grand, J. (1990). Equity versus efficiency: the elusive trade-off. *Ethics*, 100(3):554–568.

Mason, E. (2011). Value pluralism. In: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2011 Edition, [Online available at: `http://plato.stanford.edu/archives/fall2011/entries/value-pluralism/`]

MasColell, A., Whinston, M.D., and Green, J. (1995). *Microeconomic Theory.* Oxford University Press, Oxford.

Nagel, T. (1979). The fragmentation of value. In: Nagel, T. (ed.), *Mortal Questions*, Cambridge University Press, Cambridge. 128–141.

Okun, A. (1975). *Equality and Efficiency: the Big Trade-Off.* Brookings Institution, Washington.

Pattanaik, P. and Xu, Y. (2012). On dominance and context-dependence in decisions involving multiple attributes. *Economics and Philosophy*, 28(2):117-132.

Putterman, L. Roemer, J., and Silvestre, J. (1998). Does egalitarianism have a future? *Journal of Economic Literature*, 36(2):861-902.

Rabinowicz, W. (2008). Value relations. *Theoria*, 74(1):18–49.

Rawls, J. (1971). *A Theory of Justice.* Harvard University Press, Cambridge, Mass.

Rawls, J. (1987). Basic liberties and their priority. In: McMurrin, S. (ed.), *Liberty, equality and law: Selected Tanner lectures on moral philosophy*, Cambridge University Press, Cambridge. 1–87.

Raz, J. (1997). Incommensurability and agency. In: Chang, R. (ed.), *Incommensurability, Incomparability and Practical Reason*, Harvard University Press, Cambridge, Mass. 1–34.

Raz, J. (1999). *Engaging Reason: On the Theory of Value and Action.* Oxford University Press, Oxford.

Roemer, J. (1996). *Theories of Distributive Justice.* Harvard University Press, Cambridge, Mass.

Scanlon, T. (1998). *What We Owe Eachother.* Belknap Press of Harvard University Press, Cambridge, Mass.

Sen, A. (1970a). The impossibility of a paretian liberal. *Journal of Political Economy*, 78(1):152–157.

Sen, A. (1970b). *Collective Choice and Social Welfare.* Holden-Day, San Francisco.

Sen, A. (1977). Rational fools: a critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, 6(4):317–344.

Sen, A. (1979a). Personal utilities and public judgements: or what's wrong with welfare economics. *Economic Journal*, 89:537–558.

Sen, A. (1979b). Utilitarianism and welfarism. *Journal of Philosophy*, 76(9):463–489.

Sen, A. (1980). Equality of what? In: McMurrin, S. (ed), *The Tanner Lecture on Human Values*, Vol. 1, Cambridge University Press, Cambridge. 197-220.

Sen, A. (1985). Well-Being, agency and freedom: the Dewey Lectures 1984. *Journal of Philosophy*, 82(4):169–221.

Sen, A. (1997). Maximization and the act of choice. *Econometrica*, 65(4):745–779.

Sen, A. (1999). The possibility of social choice. *American Economic Review*, 89(3):349–378.

Sen, A. (2000). Consequential evaluation and practical reason. *Journal of Philosophy*, 97(9):477–502.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review.* Cambridge University Press, Cambridge.

Sinnott-Armstrong, W. (2012). Consequentialism. In: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2012 Edition, [Online available at: `http://plato.stanford.edu/archives/win2012/entries/consequentialism/`]

Suzumura, K. (1987). Social welfare function. In: Eatwell, J., Milgate, M. and Newman, P. (eds), *The New Palgrave: A Dictionary of Economics*, Vol. 4, Macmillan, London. 418–420.

Walker, M. (1977). On the existence of maximal elements. *Journal of Economic Theory*, 16:470–474.

Williams, B. (1981). *Moral Luck: Philosophical Papers.* Cambridge University Press, Cambridge.

World Bank (2012). *Inclusive Green Growth: The Pathway to Sustainable Development.* World Bank, Washington, D.C.

Zimmerman, M. (2010). Intrinsic vs. extrinsic Value. In: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2010 Edition, [Online available at: `http://plato.stanford.edu/archives/win2010/entries/value-intrinsic-extrinsic/`]

Appendix

A.1 Proof of Proposition 2

For finite sets, the proof is based on Sen (1970b: 30). Define $\succeq_e$ as the binary relation that orders states according to all values $v \in V$. So, $y^a \succeq_e y^b \leftrightarrow [y^a \succeq_v y^b$ for all $v \in V]$.

First, show that $\succeq_e$ is a quasi-ordering as it is reflexive and transitive. Reflexivity is obvious as for all $y \in Y$ one has $y \succeq_e y$. Transitivity can be shown as follows: If $y^a, y^b, y^c \in Y$ and $y^a \succeq_e y^b$ and $y^b \succeq_e y^c$, yields $y^a \succeq_v y^b$ and $y^b \succeq_v y^c$ for all $v \in V$. Due to transitivity of each value relation $\succeq_v$ this leads to $y^a \succeq_v y^c$ for all $v \in V$ and, thus, $y^a \succeq_e y^c$.

Second, show that any finite quasi-ordered set Y has at least one value-efficient state. This can be shown for a finite set of states $Y = \{y^1, y^2, ..., y^z\}$. Using the recursive rule that: $b^{i+1} = y^{i+1}$ if $y^{i+1} \succeq_e b^i$ and otherwise $b^{i+1} = b^i$. Starting at $b^1 = y^1$ one arrives after a finite number of steps at b^z which represents a value-efficient state.

For infinite sets there exists at least one value-efficient state y in Y, if there exists at least a maximal state for each $\succeq_v$ with $v \in V$. Conditions (i) and (ii) give conditions that ensure the existence of a maximal state for a respective value relation $\succeq_v$. The existence of at least one value-efficient state y requires that every $\succeq_v$ satisfies either (i) or (ii):

(i): Assume value relation $\succeq_v$ with $v \in V$ satisfies Assumption 1 and continuity, then there exists a continuous index function according to Proposition 1. As defined in Section 1.3 the set of states Y is non-empty, closed and bounded, and therefore, compact. By the Bolzano-Weierstrass Theorem there exists a maximum state y for a continuous index function I_v with $v \in V$ on a compact set. Since every maximum state is also a maximal state, this is shows the existence of a maximal state for $\succeq_v$ under Assumption 1 and continuity.

(ii): This proof follows the β-procedure in Houy and Tadenuma (2009) which has been discussed for different lexicographic relations. In our definition $\succeq_v$ is lexicographic for two attributes, if there exist two attributes $y_i \in \{y_1, ..., y_m\}$ and $y_j \in \{y_1, ..., y_m\}$ for which $y^a \succeq_v y^b$ if and only if $y_i^a > y_i^b$ or $y_i^a = y_i^b$ and $y_j^a \geq y_j^b$.

The "greater-than" ordering $>$ is defined on the subset of y_i-attributes which contains only real numbers due to $Y \subseteq \mathbb{R}^m$ and is transitive and asymmetric and, therefore, a strict partial order. Due to the compactness of the subset of y_i-attributes and the strict partial order of $>$, there exists a maximum real number y_i^* for the subset of y_i-attributes. Further, there must exist at least one state $y \in Y$ that contains the maximum real number y_i^*.

Next, construct the subset of states that all share the maximum real number for the y_i-attribute $Y^i = \{y \in Y | y_i = y_i^*\}$. In the subset Y^i consider the subset of y_j-attributes which again contains real numbers due to $Y \subseteq \mathbb{R}^m$. The "greater-than-or-equal" ordering $\geq$ is defined on the subset of y_j-attributes and is reflexive, transitive, antisymmetric and complete and, therefore, a linear ordering. Due to the compactness of the subset of y_j-attributes and the linear ordering of $\geq$, there exists a unique real number y_j^* for the subset of y_j-attributes. Further, there must exist at least one state $y \in Y$ that contains the maximum real number y_j^*.

Thus, there exists at least one state $y_v^* = (y_1, ..., y_i^*, ..., y_j^*, ..., y_m)$ in Y which contains the maximum real numbers y_j^* and y_i^* which ensures existence of a maximal state for a lexicographic relation $\succeq_v$ on the compact set Y.

A.2 Proof of Proposition 3

(i): Consider two value-efficient states of affairs $y, y' \in Y$ and a value set with $\#V = l$ values. For two values $\#V = l = 2$ and $v, w \in V$, value-efficiency of y, y' requires a trade-off between v and w, i.e. $y \prec_v y'$, $y \succ_w y'$. For more than two values $\#V = l > 2$, value-efficiency of y, y' requires a trade-off between at least two values. To show this assume there exists no trade-off between any two values in y, i.e. $y \prec_v y'$ or $y \sim_v y'$ for all $v \in V$, then y' is not value-efficient as Definition 9 that there must exists no other state of affairs $y' \in Y$ for which $y' \succeq_v y$ for all $v \in V$ and $y' \succ_v y$ for at least one $v \in V$.

(ii): Take two values v, w from a value set V with $\#V = l$ values and any two states of affairs $y, y' \in Y$. If $y \prec_v y'$ does necessarily lead to $y \succ_w y'$ for some $w \in V$, then y is a value-efficient state of affairs as Definition 9 says there must exist no $y' \in Y$ for which $y \prec_v y'$ for some $v \in V$ and $y \preceq_w y'$ for all $w \in V$.

(iii): Assume there are 3 values in the value set $u, v, w \in V$ and two states of affairs $y, y' \in Y$. Assume the orderings necessary for a win-win relationship between u and v: $y \prec_u y'$, $y \prec_v y'$; and show that $y, y' \in Y$ can still be value-efficient. Assume $y \succ_w y'$ so that y' does not dominate y. This yields two value-efficient states of affairs y, y' where y exhibits a win-win relationship between $u, v \in V$.

To show an independence relationship in a value-efficient state of affairs assume one further state of affairs y''. Assume additionally the orderings necessary for an independence relationship between u and v: $y \prec_u y''$, $y \sim_v y''$; and show that $y, y', y'' \in Y$ can still all be value-efficient. Assume $y \succ_w y'$, $y \succ_w y''$ so that y' and y'' do not dominate y. Assume $y' \sim_w y''$ and $y' \prec_u y''$, $y' \succ_v y''$ so that y' does not dominate y''. This yields three value-efficient states of affairs y, y', y'' where y exhibits an independence relationship between $u, v \in V$.

A.3 Proof of Proposition 4

If all $v \in V$ are individual preferences, then all value relations are identical to preference relations. This makes value-efficiency equivalent to Pareto-efficiency.

A.4 Proof of Corollary 1

Trade-off relationship

Assume there are 4 values in the value set: two individual preferences $p_1, p_2 \in P \subset V$ and two non-preference values $v, w \in V$ and two states of affairs y, y'. Assume the orderings necessary for both states to be Pareto-efficient: for the first individual $y \prec_{p_1} y'$, and for the second individual $y \succ_{p_2} y'$. Further, assume orderings for a trade-off between v and w: $y \prec_v y'$ and $y \succ_w y'$. This yields two Pareto-efficient and value-efficient states of affairs y, y' where y exhibits a trade-off relationship between $v, w \in V$.

Win-win relationship

Assume there are 4 values in the value set: two individual preferences $1, 2 \in P \subset V$ and two non-preference values $v, w \in V$ and two states of affairs y, y'. Assume the orderings necessary for both states to be Pareto-efficient: for the first individual $y \prec_{p_1} y'$, and for the second individual $y \succ_{p_2} y'$. Further, assume orderings for a win-win relationship between v and w: $y \prec_v y'$ and $y \prec_w y'$. This yields two Pareto-efficient and value-efficient states of affairs y, y' where y exhibits a trade-off relationship between $v, w \in V$.

Independence relationship

Assume there are 4 values in the value set: two individual preferences $1, 2 \in P \subset V$ and two non-preference values $v, w \in V$ and three states of affairs y, y', y''. Assume the orderings necessary for all three states to be Pareto-efficient: for the first individual $y \prec_{p_1} y'$, $y \prec_{p_1} y''$, $y' \prec_{p_1} y''$, and for the second individual $y \succ_{p_2} y'$, $y \succ_{p_2} y''$, $y' \succ_{p_2} y''$. Further, assume orderings for an independence relationship between v and w: $y \prec_v y'$, $y \prec_v y''$ and $y \prec_w y'$, $y \sim_w y''$. Assume $y' \sim_v y''$ and $y' \prec_w y''$, so that y' does not dominate y''. This yields three Pareto-efficient and value-efficient states of affairs y, y', y'' where y exhibits an independence relationship between $v, w \in V$.

A.5 Proof of Proposition 5

Assume a value set with three values: two individual preferences $p_1, p_2 \in P \subset V$ and a non-preference values $v \in V$ and two states of affairs y, y'. Then it can be that $y \prec_{p_1} y'$ and $y \prec_{p_2} y'$, but $y \succ_v y'$. This yields two value-efficient states of affairs y, y', where state y is not Pareto-efficient.

Paper 2:

Economic analysis of trade-offs between justices

Economic analysis of trade-offs between justices

Stefan Baumgärtner[a], Stefanie Glotzbach[a], Nikolai Hoberg[a], Martin F. Quaas[b], Klara Helene Stumpf[a] *

[a] Department of Sustainability Sciences and Department of Economics,
Leuphana University of Lüneburg, Germany

[b] Department of Economics, University of Kiel, Germany

Abstract: We argue that economics - as the scientific method of analysing trade-offs - can be helpful (and may even be indispensable) for assessing the trade-offs between intergenerational and intragenerational justice. Economic analysis can delineate the "opportunity set" of politics with respect to the two normative objectives of inter- and intragenerational justice, i.e. it can describe which outcomes are feasible in achieving the two objectives in a given context, and which are not. It can distinguish efficient from inefficient uses of instruments of justice. It can identify the "opportunity cost" of attaining one justice to a higher degree, in terms of less achievement of the other. We find that, under very general conditions, (1) efficiency in the use of instruments of justice implies that there is rivalry between the two justices and the opportunity cost of either justice is positive; (2) negative opportunity costs of achieving one justice exist if there is facilitation between the two justices, which can only happen if instruments of justice are used inefficiently; (3) opportunity costs of achieving one justice are zero if the two justices are independent of each other, which is the case in the interior of the opportunity set where instruments of justice are used inefficiently.

Correspondence: Stefan Baumgärtner, Leuphana University of Lüneburg, Sustainability Economics Group, P.O. Box 2440, D-21314 Lüneburg, Germany, phone: +49.4131.677-2600, fax: +49.4131.677-1381, email: baumgaertner@uni.leuphana.de.

* We thank Joachim Fünfgelt, Michael Schefczyk, the editors and two anonymous reviewers for helpful comments. Financial support from the German Federal Ministry of Education and Research (BMBF) under grants 01UN1011A/B is gratefully acknowledged.

This paper is published in *Intergenerational Justice Review* 1/2012, 4-9 and has been reformatted for this thesis. A first draft of this paper became available in University of Lüneburg Working Paper Series in Economics (No. 218) in November 2011.

2.1 Introduction

Justice is a multifarious normative idea about the quality of relationships among members of society. One may argue that there are many "justices", insofar as different parts of society, different types of relationships, or different substantive areas are addressed. The overall societal goal ("vision") of sustainability particularly addresses two justices: (i) justice between presently living persons ("intragenerational justice"), and (ii) justice between members of present and future generations ("intergenerational justice").[1,2]

With two (or more) different justices as normative objectives of equal rank, it may be that there exists a trade-off between them, that is, performing better with regard to one objective implies performing worse with regard to the other one. In particular, it may be that fostering intragenerational justice makes it more difficult to attain intergenerational justice, and vice versa. Such a trade-off at the level of normative objectives of equal rank - if it exists - asks for societal resolution. The question is: How to act in the face of different justices? Important examples for such a trade-off include government spending on social welfare vs. investment in public infrastructure and education, or the exploitation vs. conservation of non-renewable natural resources.

In this essay, we argue that economics - as the scientific method of analyzing trade-offs - can be helpful (and may even be indispensable) for assessing the trade-offs between different justices. We understand economics as being defined by its method, rather than by its substance matter or by some normative objective[3], and we sketch how to employ this method to analyse trade-offs between justices. An important contribution that economics can make to this analysis is to introduce the secondary normative criterion of efficiency which characterises the non-wasteful use of scarce resources to attain the primary normative objectives of justice: a situation is efficient with regard to different objectives if it is not possible to improve on one objective without doing worse on another

[1] WCED 1987.

[2] In addition, some conceptions of sustainability also include justice towards nature as a third normative objective of equal rank.

[3] This is the standard interpretation of modern economics according to Robbins 1932. For an encompassing discussion of this and other interpretations of economics, see Hausman 2007.

one. Being derived from primary normative objectives, the criterion of efficiency itself makes a normative claim: it is good to use scarce resources efficiently to attain intra- and intergenerational justice; it is wrong to use scarce resources inefficiently for that purpose.

This approach of using economics as a method to study the efficient use of scarce resources in the attainment of rivaling normative objectives of justice[4] opens an innovative perspective on what the role of economics should be (as a method) in the discussion of justice, and on how to bridge the gap - systematically and rigorously - between ideal theory and non-ideal politics.

2.2 Specifying justice(s)

To inform our understanding of intra- and intergenerational justice, the abstract and general concept of justice needs to be further specified. We take justice to generally refer to the mutual claims of members of the community of justice from the standpoint of impartiality.[5] This minimum definition leaves ample room for very different, and sometimes much contested, conceptions of justice. Each of them can be described more precisely by specifying a number of elements in a "syntax of justice".[6,7]

The community of justice. Justice refers to mutual claims[8] within a community of justice. We term those holding a particular claim the claim holders, and those re-

[4] This approach, as applied to the three justices included in the vision of sustainability - intra- and intergenerational justice as well as justice towards nature - has been called "sustainability economics" (Baumgärtner and Quaas 2010, Baumgärtner 2011).

[5] E.g. Gosepath 2007: 82.

[6] Baumgärtner / Glotzbach / Stumpf 2011. This "syntax" is our approach to structure what has been called the different "dimensions" (Pogge 2006, Dobson 1998, see also Ott and Döring 2008) of the concept of justice. It allows fully specifying a particular conception of justice.

[7] In the following, we specify the essential elements of the syntax to clarify the conceptions of inter- and intragenerational justice.

[8] Young 1994, Ott and Döring 2008: 59 et seqq.

sponsible for the fulfillment of the claim the claim addressees.[9] Intragenerational justice entails claims held by currently living persons (claim holders) towards other currently living persons (claim addressees). Intergenerational justice entails claims held by persons living in the future ("future generations", claim holders) towards persons living today (claim addressees).[10] It is not necessary that such a claim is explicitly put forward by the claim holder (which may be impossible in the case of intergenerational justice). What matters is that a legitimate claim might be formulated by someone speaking for the claim holder.

Positive and negative claims. Generally, claims can be positive, i.e. defining an entitlement to a certain good,[11] or negative, i.e. demanding freedom from harm.[12] Claims are considered legitimate if they could be agreed on from the standpoint of impartiality and equal consideration. For example, intergenerational justice claims could be specified as a positive claim of future generations to certain stocks and systems, such as a democratic political system, a stock of manufactured capital and critical knowledge, or intact ecosystems, implying a responsibility of the present generation to pass on these stocks and systems in a good state to future generations. Future generations may also have a negative claim: not to be harmed by any activities of the presently living generation, e.g. through increasing systemic risks caused by a dysfunctional global financial system or through nuclear waste left over as a by-product of present electricity

[9] The delineation of the community of justice, especially the question of who is to be included as a claim holder, can be drawn according to different criteria such as reciprocity, dignity, ability to experience pain, etc. (e.g. Baumgärtner, Glotzbach and Stumpf 2011).

[10] The third justice often included in sustainability conceptions, justice towards nature, refers to claims held by "nature", e.g. higher non-human animals capable of experiencing pain or of pursuing goals, against humanity. Thus, the claim holders differ, while the claim addressees belong to the group of currently living persons in all three cases. While intra- and intergenerational justice reflect an anthropocentric idea of justice, according to which nature matters to humans exclusively because of its instrumental value, the idea of justice towards nature assigns an intrinsic value to nature (Baumgärtner and Quaas 2010: Sec. 2), so that "nature" becomes a claim holder in its own right.

[11] "Goods" should be understood in a wide sense.

[12] cf. Baumgärtner / Glotzbach / Stumpf 2011.

production. Intragenerational justice claims include the positive claim for satisfaction of basic needs, and the negative claim that one's freedoms should not be harmed (human rights).

Judicandum. We use the term judicandum to describe that which is to be judged as just or unjust. Judicanda can be agents, actions, institutions or states of the world.[13] When discussing inter- and intragenerational justice, the judicanda could be the actions of currently living persons (and the consequences of these actions, such as, say, the distribution of certain primary goods), as the claim addressees of both justices belong to the current generation.

Instruments of justice. We use the term instrument of justice to describe that which is to be used to satisfy the legitimate claims of justice. In many conceptions of justice, these will be objects of distribution (answers to the question "What is distributed?"[14]), but the satisfaction of legitimate claims could also be achieved via, say, institutional reform to ensure procedural justice. So, the question here is how legitimate claims are addressed. For example, one instrument of intergenerational justice could be the investment in public goods such as education and infrastructure, or the distribution of stocks of non-renewable resources between different generations. The aim of intragenerational justice could, for example, require institutional reform of international trade rules ("fairness").

Metric for the judgment. For statements about the degree of attainment of a normative objective, there must be some way to measure the justice of the judicanda: one needs a metric to judge whether, and to what extent, a judicandum is just or unjust. For this metric, different informational bases have been proposed, such as e.g. capabilities, primary goods, or utility.[15] It is possible to use different metrics for inter- and intragenerational justice.

In sum, judging a certain judicandum as inter- or intragenerationally just according to a metric requires first to specify the positive and negative claims of claim holders in

[13] Pogge 2006: 863.

[14] Sensu Dobson 1998: 73 et seqq.

[15] Cf. Pogge 2006: 868.

present and future generations against claim addressees in the present generation, which are to be satisfied by certain instruments of justice.

As we discuss two different justices, both of which demand the fulfillment of legitimate claims through the use of instruments of justice by the same addressee, a non-trivial decision problem arises for this addressee - the present generation. We therefore need to have a closer look at the possible relationships of these two justices.

2.3 Relationships between justices

Generally, the two justices are related both on the "value" side and the "production" side.[16] On the value side, the relationship refers to the desirability, from a societal point of view, of attaining one justice relative to the other one. For example, society may be willing to trade-off one justice against the other[17], or one justice might strictly dominate the other. In this essay, we build on the minimal and very general premise, widely held in the literature,[18] that both intra- and intergenerational justice are considered by society as desirable normative objectives of equal rank. Beyond that, we do not further discuss the value side.

On the production side, the relationship refers to the feasible outcomes of the use of instruments of justice, that is, combinations of degrees of attainment of both justices. Here, what is feasible is determined by the structure and functioning of the given system, based on natural resource endowments, technology, institutions, etc. The set of all feasible combinations in terms of the two justices is called the "opportunity set". It describes society's options for choice, which are independent of what society considers desirable. That is, the production side and the value side are independent of each other.

Scientific analysis and political implementation have shown that, in general, three relationships may hold on the production side between intra- and intergenerational jus-

[16] LeGrand 1990: 555.

[17] Barry 1965: Sec. 1.

[18] E.g. Dobson 1998: 3 et seqq., Ott / Döring: 2008, Visser´t Hooft 2007: 56, WCED 1987: 43.

tice:[19]

- (1) Independency: The objectives of intra- and intergenerational justice can be achieved independently, that is, attaining one objective to a higher degree does not necessitate any change in the degree to which one attains the other one.[20]
- (2) Facilitation: Achieving one objective supports achieving the other one, that is, attaining one objective to a higher degree induces a higher degree of attainment of the other one.[21,22]
- (3) Rivalry: A fundamental rivalry (or "trade-off") exists between the objectives of intra- and intergenerational justice, that is, attaining one objective to a higher degree necessarily reduces the degree to which one attains the other one.[23]

For illustration, we give examples from different contexts. Independency is an assumption frequently made in ecological, environmental and resource economics.[24] For example, cap-and-trade systems for greenhouse gas emissions imply that the overall intergenerational impact on global climate can be governed independently of the initial intragenerational distribution of emission certificates.[25] Facilitation is prominently stated with regard to the provision of public goods. For instance, public investment in education or the improvement of public transportation systems may simultaneously

[19] Here, we extend the argument from Glotzbach and Baumgärtner (in press, Sec. 3) which originally refers to justice with regard to the use and conservation of ecosystems.

[20] Independency does not need to be symmetric: achieving one objective may be independent of achieving the other one, but not vice versa.

[21] This relationship is similar to the concept of "joint production" in economics, which means that the production of a wanted good necessarily gives rise to additional outputs (cf. Baumgärtner et al. 2006).

[22] This facilitation may be one-way, or the other way, or a mutual facilitation between the achievement of the two objectives.

[23] Like independency and facilitation, rivalry does not need to be symmetric.

[24] E.g. Dasgupta and Heal: 1979.

[25] E.g. Perman et al: 2003: 219 et seqq.

benefit today's poor and future persons. Rivalry is often assumed when the possibility of intragenerational redistribution of access rights to rival resources is heavily limited. In such cases, meeting the legitimate claims of the poor to the resource possibly reduces the total resource stock passed on to future generations and, thereby, may be at the expense of intergenerational justice. For example, if the government spends a higher share of tax revenue to increase social support of the poor without being able to enforce higher taxes on the rich, the government has less revenue to invest in public infrastructure and education.

A host of specific determinants - natural, technological and institutional factors - impact on the production relationship between intra- and intergenerational justice, for example because they influence the availability and effectiveness of the instruments of justice. Thereby, they affect which relationship holds. Two examples for such determinants are population development and political restrictions. In many countries of the global North, a population development characterised by higher life expectancy and lower birth rates challenges the existing social security systems. A potential trade-off among the goal to reduce old-age poverty (intragenerational justice), and the goal to avoid an unacceptable high financial burden on the young generation (intergenerational justice) may occur. Political restrictions limit the political scope for redistribution of resources within a society. If, for instance, the political scope for redistribution of wealth within a society is tight due to resistance against introduction of an inheritance tax, the situation of the poor can only be improved by increasing public expenditures and, thereby, possibly adding to public debt in the long-term - therefore causing a trade-off between inter- and intragenerational justice.

Regarding the production relationship between intra- and intergenerational justice in the use and conservation of ecosystem services, Glotzbach and Baumgärtner (in press: Sec. 4) found that the determinants impacting on this relationship are the quantity and quality of ecosystem services, population development, the substitutability of ecosystem services by human-made goods and services, technological progress, and institutions and political restrictions. The determinant substitutability of ecosystem services, for instance, influences the character of the relationship between the justices as follows: if an ecosystem service is substitutable by human made goods and services, an overexploitation of the ecosystem service by members of the present generation to increase intragenerational justice can be compensated by sufficient investment in other forms of physical, social and human capital to secure intergenerational justice - the relationship between the justices is one of independency or facilitation. If an ecosystem service is non-substitutable, an overexploitation of the ecosystem service by members of the present generation to increase intragenerational justice cannot be compensated and, hence, reduces the degree of intergenerational justice - the relationship between the justices is one of rivalry.

In sum, the opportunity set, which embodies information on the production relationships between the two justices in all feasible outcomes, crucially depends on a number of fundamental context-specific determinants.

2.4 Scarcity, economic efficiency, and opportunity costs

Irrespective of which production relationship holds between inter- and intragenerational justice, society has to make a decision on how to use some instruments of justice in the attainment of these objectives. Very often, the use of instruments of justice means employing scarce resources that may be used in alternative ways.[26] This is where the key contribution of economics to the study of societal problems comes in: How to use scarce resources efficiently in the attainment of some objectives? According to a classical definition, economics

> studies human behaviour as a relationship between [given] ends and scarce means which have alternative uses.[27]

[26] Scarcity is generally considered as central to many important problems of justice (Dobson 1998: 12).

[27] Robbins 1932: 15.

With this definition, economists generally understand efficiency as non-wastefulness in the use of "scarce means" to attain some "ends" that humans pursue in their actions. In this understanding, ends are open-ended: they are not determined by economics as a method. In principle, it could be any ends that humans pursue. Here, we focus on intra- and intergenerational justice as two primary normative objectives that humans pursue.[28] Then, drawing on the common definition of efficiency by Pareto (1906),[29] one can define efficiency as follows:

> An allocation of resources is efficient if it is impossible to move toward the attainment of one social objective without moving away from the attainment of another objective.[30]

The minimal assumption needed to define efficiency in this way is that, for each justice, the metric of justice allows a distinction to be made between a higher and a lower degree of attainment of the respective justice. In particular, it is neither necessary to assume cardinality of each metric nor commensurability of the two justices.[31] Thus, this notion of efficiency and the subsequent analysis are very general.

If efficiency is related in this manner to some primary normative objectives, it acquires the status of a secondary normative objective.[32,33] This means, it is good to

[28] This goes beyond what economists usually consider as ends (cf. Baumgärtner 2011). Traditionally, economics has been concerned with the end of an ever better satisfaction of human needs and wants. This end can be further specified and operationalised as individual utilities (microeconomics), or as policy goals such as low inflation and low unemployment (macroeconomics).

[29] According to the original criterion of Pareto (1906), which assesses allocations based on the well-being of individual persons, an allocation of resources is efficient if no one can be made better off (in terms of this person's individual utility) without making anyone else worse off (in terms of the other person's individual utility).

[30] LeGrand 1990: 559.

[31] A cardinal metric is one that preserves orderings uniquely up to linear transformations; commensurability of justices means that the metric of both justices is in the same units.

[32] LeGrand 1990: 560.

[33] Here, we study the relationship, including a potential trade-off, between two primary normative objectives. There is also a discussion on the so-called "equity-efficiency trade-off" (surveyed by e.g.

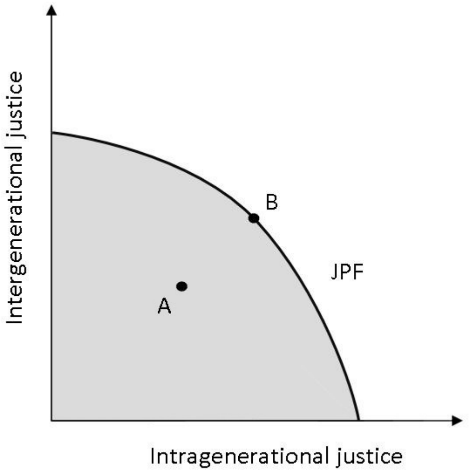

Figure 2.5: Rivalry and independency

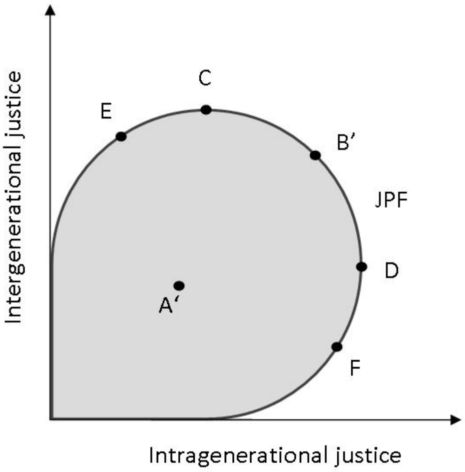

Figure 2.6: Rivalry, facilitation, and independency

use resources efficiently; it is wrong to use them inefficiently. In this perspective, the contribution of economics to the study of societal problems lies in characterising the (in)efficient use of scarce means in the attainment of multiple primary normative objectives. For this purpose, economics provides a broad set of methods to analyse, display and empirically verify the relationships between these objectives.

Figures 2.5 and 2.6 illustrate the opportunity set and efficiency in attaining the two normative objectives of intra- and intergenerational justice. The axes indicate the degree of attainment of inter- and intragenerational justice, respectively, based on the respective metrics of justice. Thus, each point in the diagram represents an outcome of the use of the instruments of justice. In Figure 2.5, the shaded area depicts all feasible outcomes in the given context, that is, for given resource endowment, technology, institutions, and the like ("opportunity set"). The curve JPF ("justice possibility frontier") denotes its frontier. Outcomes to the northeast of this curve are not feasible in the given

Putterman et al. 1998), where equity and efficiency are treated as normative objectives of equal rank. But efficiency - in contrast to equity - cannot serve as a primary normative objective, so that this trade-off is irrelevant (LeGrand 1990: 566).

context. Point A represents an outcome where the instruments of justice are used in an inefficient manner as more intergenerational justice could be achieved without sacrificing intragenerational justice. In contrast, the use of the instruments of justice in point B is efficient as no higher degree of attainment of one justice is feasible without reducing the other one. Generally, all outcomes below the JPF-curve correspond to inefficient uses of the instruments of justice, whereas all outcomes on the curve correspond to efficient uses of these instruments.

Obviously, in point B there is rivalry between intragenerational and intergenerational justice: attaining one to a higher degree necessarily reduces the degree to which one attains the other one. This loss can be measured by the concept of "opportunity cost". The opportunity cost of increasing, say, intragenerational justice is the corresponding minimal loss of intergenerational justice. In contrast, in point A there is independency between intragenerational and intergenerational justice: attaining one to a higher degree does not necessitate any change in the degree to which one attains the other one. Hence, there are no opportunity costs of increasing one or the other justice. Generally, in all efficient outcomes, i.e. on the JPF-curve, there is rivalry between the two justices and, thus, positive opportunity costs. In all inefficient outcomes, i.e. under the JPF-curve, there is independency between the two justices and, thus, zero opportunity costs.

For example, the opportunity set of Figure 2.5 may refer to the use of a non-renewable natural resource such as oil or gas: the resource may be exploited today for social welfare policy (intragenerational justice); alternatively, it may be conserved for future generations (intergenerational justice).

In a different context, the opportunity set may look as in Figure 2.6. The shaded area again depicts all outcomes that are feasible in this context ("opportunity set"), with the JPF-curve as its frontier. As in Figure 2.5, outcomes A' and B' correspond to an inefficient and an efficient use, respectively, of the instruments of justice. Obviously, all points on the JPF-curve between C and D represent outcomes of efficient uses of the instrument of justice, because no higher degree of attainment of one justice is feasible without reducing the other one. These outcomes are characterised by rivalry between the two justices and positive opportunity costs of either justice.

Outcome E is inefficient, but as it lies on the JPF, attaining intergenerational justice to a higher degree starting from this point necessarily also leads to a higher degree of intragenerational justice. That is, in outcome E there is facilitation between the two justices. But facilitation is not symmetric: attaining a higher degree of intragenerational justice, starting again from point E, does not necessarily induce a higher degree of intergenerational justice. Hence, the opportunity cost of increasing intergenerational justice is negative: increasing intergenerational justice does not incur a loss, but a gain, of intragenerational justice, and the opportunity cost of increasing intragenerational justice is zero. In outcome F, the situation is reversed: attaining intragenerational justice to a higher degree facilitates attaining intergenerational justice to a higher degree, but not vice versa; hence, the opportunity cost of increasing intragenerational justice is negative, while the opportunity cost of increasing intergenerational justice is zero. Generally, all (inefficient) uses of instruments of justice along increasing parts of the JPF correspond to outcomes where attaining one justice to a higher degree facilitates attaining the other one, but not vice versa, so that the former has negative opportunity cost, while the latter has zero opportunity cost.

For example, the opportunity set of Figure 2.6 may refer to government spending on education, where a broader educational base decreases income inequality within a generation (intragenerational justice), and at the same time increases prospects for economic growth over time (intergenerational justice).

As the figures and examples illustrate, the shape of the opportunity set may differ from context to context, and with it the relationships between the two justices.[34] As the opportunity set is fundamentally determined by natural resource endowment, technol-

[34] In addition to the two fundamental shapes of the opportunity set discussed here, other shapes are imaginable. For example, the justice possibility frontier may be linearly downward sloping, implying constant opportunity costs in all efficient outcomes. It may also be convex (resulting e.g. from increasing returns to scale in the use of instruments of justice), and the frontier may not even intersect but asymptotically approach the axes. This would imply that the opportunity costs of one justice may rise to infinity. Yet, all insights into the relationships between the two justices and efficiency that are essential for our main line of argument can already be obtained from the two shapes of the opportunity set presented here. We therefore refrain from discussing additional shapes in detail.

ogy, institutions, etc. (cf. Section 3), a change in these fundamental determinants may change the opportunity set and the relationships between the two justices. For example, with given endowment of a non-renewable resource, technical progress in resource extraction would shift the JPF-curve in Figure 2.5 outwards.

2.5 Conclusion

Robbins' (1932) definition of economics delimits the contribution of economics to the study of normative questions. It does not lie in determining what ends to pursue or in developing the means to achieve a normative objective. Rather, the focus of economic analysis is on efficiency, i.e. non-wastefulness in the use of scarce resources that have alternative uses as means to attain given normative objectives. Thus, in contexts where there is no scarcity or no alternatives exist, economics does not lend itself to the discussion of normative questions. Yet, many questions of justice arise under conditions of scarcity and involve the freedom to make choices. Such questions can be discussed in economic terms.

Economic analysis of inter- and intragenerational justice builds on three fundamental, and rather weak, assumptions:

(1) On the "value" side, the two justices are considered by society to be of equal rank.

(2) For each justice, one can measure the degree to which one attains this justice. This measurement does not need to be cardinal but may be ordinal, and the two justices do not need to be commensurable but the two metrics may be in different units.

(3) For a given context - specified by natural, technological, institutional factors, etc. - one can describe the outcome of using scarce resources (as instruments of justice) in terms of these measures of the two justices.

With these assumptions, the genuine and original contribution of an economic analysis of justice is threefold:

(1) Economic analysis can delineate the "opportunity set" of politics with respect to the two normative objectives of inter- and intragenerational justice, i.e. it can describe which outcomes are feasible in achieving the two objectives in a given context, and which are not. The opportunity set includes information on whether the production relationship between the two justices in some outcome is one of rivalry (i.e. trade-off), independency, or facilitation; and it distinguishes efficient from inefficient allocations of scarce resources.

    As efficiency, when related to the primary normative objectives of intergenerational and intragenerational justice, is a secondary normative objective, one conclusion for policy-making is straightforward: instruments of justice should be used efficiently; they should not be used inefficiently.

    One important conclusion about the production relationship between intra- and intergenerational justice follows directly from the very definition of efficiency. In outcomes of efficient resource use there is always rivalry between the different justices - attaining one justice to a higher degree necessarily reduces the degree to which the other is attained. In contrast, in outcomes of inefficient resource use there is either independency between the two justices - the level of attainment of one justice can be improved without doing worse on the other one, or even both can be improved - or facilitation - improving the level of attainment of one justice necessarily also improves the other one.[35]

(2) Based on the opportunity set, economic analysis can identify the "opportunity cost" of attaining one justice to a higher degree, in terms of less achievement of the other. Positive opportunity costs of achieving one justice exist if there is rivalry between the two normative objectives of intergenerational and intragenerational justice; negative opportunity costs of achieving one justice exist if there is facilitation between the two justices; opportunity costs are zero if there is independency between the two justices. Generally, negative and zero opportunity costs indicate

[35] In the (inefficient) interior of the opportunity set there is always independency; and facilitation can only occur on the inefficient part of the justice possibility frontier.

inefficiency in the allocation of resources, while positive opportunity costs indicate an efficient resource allocation.

(3) Economic analysis can identify how the opportunity set changes as its determinants - natural, technological, institutional factors, etc. - change. In particular, it can study how the occurrence and extent of rivalry, independency or facilitation in the relationship between the two justices changes as underlying determinants change. Hence, it may suggest how to manage these underlying determinants in order to decrease the degree of rivalry and to increase the degree of independency or facilitation.

The economic analysis presented here cannot determine which of the efficient outcomes on the justice possibility frontier is preferable. Moving from one efficient outcome to another means incurring opportunity costs - i.e. furthering the degree of attainment of one normative objective at the cost of the other one. Depending on how the relationship between the two normative objectives is shaped on the "value side", it might well be acceptable to incur these costs - for example, burdening the presently living with a small tax that would prevent future generations from huge damage.

So, economic analysis can give no clear guidance on how to decide among efficient outcomes - i.e. in the case of rivalry between objectives. Its contribution lies in pointing out clearly inefficient outcomes, and in identifying the opportunity costs of moving from one efficient outcome to another.

These insights can help make an informed decision about how to use scarce resources that have alternative uses to attain the two normative objectives of inter- and intragenerational justice in a non-wasteful manner. This seems to be a valuable contribution for societies facing decisions about the use of scarce resources in view of different normative objectives of equal rank. Of course, this would not make hard decisions easy, but at least efficiently difficult.

References

Barry, B (1965): Political Argument. New York: Humanities Press.

Baumgärtner, S (2011): Normative Begründung der Nachhaltigkeitsökonomie. In: Studierenden Initiative Greening the University e.V. (eds.): Wissenschaft für nachhaltige Entwicklung! Multiperspektivische Beiträge zu einer verantwortungsbewussten Wissenschaft. Marburg: Metropolis-Verlag, 273-298.

Baumgärtner, S / Faber, M / Schiller, J (2006): Joint Production and Responsibility in Ecological Economics. On the Foundations of Environmental Policy. Cheltenham: Edward Elgar.

Baumgärtner, S / Glotzbach, S / Stumpf, K H (2011): Nachhaltigkeit als Gerechtigkeit. Eine Einführung aus Sicht der Nachhaltigkeitsökonomie. Lecture Notes.

Baumgärtner, S / Quaas M F (2010): What is sustainability economics? In: Ecological Economics. Vol. 69 (3/2010), 445-450.

Dasgupta, P S / Heal, G M (1979): Economic Theory and Exhaustible Resources. Cambridge: Cambridge University Press.

Dobson, A (1998): Justice and the Environment. Conceptions of Environmental Sustainability and Dimensions of Social Justice. Oxford and New York: Oxford University Press.

Glotzbach, S / Baumgärtner, S (in press): The relationship between intragenerational and intergenerational ecological justice. In: Environmental Values.

Gosepath, S (2007): Gerechtigkeit. In Fuchs, D / Roller, E (eds): Lexikon Politik. Hundert Grundbegriffe. Stuttgart: Reclam, 82-85.

Hausman, D M (2007): The Philosophy of Economics. An Anthology. 3rd edition.

Cambridge: Cambridge University Press.

LeGrand, J (1990): Equity versus efficiency: the elusive trade-off. In: Ethics. Vol. 100 (3/1990), 554-568.

Ott, K / Döring, R (2008): Theorie und Praxis starker Nachhaltigkeit. 2nd edition. Marburg: Metropolis.

Pareto, V (1906): Manuale d'economia politica con una introduzione alla scienza sociale. Milano: Società editirce libraría.

Perman, R / Ma, Y / McGilvray, J et al. (2003): Natural Resource and Environmental Economics. 3rd edition. Harlow: Pearson.

Pogge, T W (2006): Justice. In: Borchert .M. (ed.): Encyclopedia of Philosophy. 2nd edition. Detroit: Macmillan Reference USA, 862-870.

Putterman, L / Roemer, J E / Silvestre, J (1998): Does egalitarianism have a future? In: Journal of Economic Literature. Vol. 36 (2/1998), 861-902.

Robbins, L (1932): An Essay on the Nature and Significance of Economic Science. London: Macmillan.

Visser't Hooft, H P (2007): Justice to Future Generations and the Environment. Berlin and New York: Springer.

[WCED] World Commission on Environment and Development (1987): Our Common Future. New York: Oxford University Press. Young, H P (1994): Equity in Theory and Practice. Princeton: Princeton University Press.

Paper 3: Irreversibility, ignorance, and the intergenerational equity-efficiency trade-off

Irreversibility, ignorance, and the intergenerational equity-efficiency trade-off

NIKOLAI HOBERG* and STEFAN BAUMGÄRTNER

Department of Sustainability Sciences and Department of Economics,
Leuphana University of Lüneburg, Germany

Abstract: Two important policy goals in intergenerational problems are Pareto-efficiency and sustainability, i.e. intergenerational equity. We demonstrate that the pursuit of these goals is subject to an intergenerational equity-efficiency trade-off. Our analysis highlights two salient characteristics of sustainability problems and policy: (i) *temporal irreversibility*, i.e. the inability to revise one's past actions; and (ii) *unawareness* of future consequences of present actions in human-environment systems ("unknown unknowns"). If initially unknown sustainability problems become apparent and policy is enacted after irreversible actions were taken, policy-making faces a fundamental trade-off between Pareto-efficiency and sustainability.

JEL-Classification: D3, H23, Q01, Q38, Q56

Keywords: climate change, closed ignorance, intergenerational equity-efficiency trade-off, irreversibility, Pareto-efficiency, sustainability, unawareness

* Corresponding author: Sustainability Economics Group, Leuphana University of Lüneburg, P.O. Box 2440, D-21314 Lüneburg, Germany, phone: +49.4131.677-2715, fax: +49.4131.677-1381, email: hoberg@uni.leuphana.de.

This is a revised version of the first draft of the paper that became available in University of Lüneburg Working Paper Series of Economics (No. 198) in February 2011.

3.1 Introduction

Global environmental indicators highlight increasing degradation in many fields such as biodiversity, climate change and non-renewable resource scarcity (e.g. MEA 2005, IPCC 2007, UNEP 2012). This trend has intensified concerns about intergenerational justice and brought the concept of sustainable development in the design of public policy (e.g. WCED 1987). For instance, advocates of "climate justice" demand the equitable distribution of the benefits and damages from CO_2-emissions between developing and industrialized countries as well as between historic and future emitters (Neumayer 2000). The challenge for sustainability policy is therefore to realize an efficient and intergenerationally just allocation of resources.[1]

Achieving *inter*generational justice and efficiency simultaneously may not always be possible as redistribution might incur an equity-efficiency trade-off. Such a trade-off is familiar from the attempt to achieve *intra*generational equity in social policy: the quest for equal utility levels (equity) incurs a (first-best) Pareto-inefficient allocation (Putterman et al. 1998, Le Grand 1990), because of different mechanisms such as incentive distortions or administrative costs. Thus, the question emerges whether policies aiming at sustainability are likewise subject to an *inter*generational equity-efficiency trade-off.

Following the intuition from the second welfare theorem, Howarth and Norgaard (1990, 1992) show that in an overlapping-generations-model both equity and efficiency can be achieved intergenerationally, given a set of public policies such as Pigouvian taxes, intergenerational transfer payments and the assignment of resource rights between generations. Krautkraemer and Batina (1999) find that a non-decreasing-utility constraint in a model with a renewable resource can lead to Pareto-inefficient overaccumulation of the resource. In that case, all generations could be made better off by allowing decreasing utility over time. Gerlagh and Keyzer (2001) compare different policy instruments for the sustainable intergenerational distribution of resources and find that a trust fund in which all natural resources and ecosystem services are administered that can be pro-

[1] Sustainability policy goes beyond mere internalization of intertemporal externalities but aims at intergenerational equity (Pezzey 2004, Baumgärtner and Quaas 2010).

duced sustainably, leads to a Pareto improvement compared to a zero-extraction policy. Considering uncertain future outcomes and preferences, Krysiak (2009) finds a trade-off between protecting future individuals from potential harm (sustainability) and thereby abstaining from actions that would have made everyone better off (efficiency).

In this paper, we investigate how an intergenerational equity-efficiency trade-off in sustainability policy emerges from the genuine character and mechanisms of intergenerational policy-making. We employ a two-non-overlapping-generations model that combines an intragenerational production decision on the use of circulating capital and a non-renewable resource, with a negative intergenerational externality. Compared to *intra*generational policy-making, there are two salient characteristics of sustainability problems and policy: (i) temporal irreversibility (Baumgärtner 2005), i.e. the inability to revise one's past actions; (ii) "closed ignorance" (Faber et al. 1992) or "unawareness" (Dekel et al. 1998), i.e. future consequences of present actions may be "unforeseen contingencies" (Dekel et al. 1998), also known as "unknown unknowns" (Rumsfeld 2002). As such unawareness is a more fundamental form of uncertainty than risk or Knightian uncertainty, common methods such as expected utility maximization or subjective probability distributions cannot be employed.

An important case in point is the current discussion of "climate justice", and here especially equity between historic and future emitters. The first generation in the model represents historic emitters (e.g. Europe and North America) who irreversibly used non-renewable fossil fuels for the production of consumption goods and, in the process, emitted greenhouse gases that lead to significant climate change. These actions were taken under initial unawareness of the effects of greenhouse gases on climate change. The second generation in the model represents future emitters (e.g. China and India) who find diminished stocks of fossil fuels and also suffer the damages from climate change. The crucial challenge now is that climate policy is being shaped and implemented after historic production and emissions have already irreversibly taken place. While the amount of fossil fuels used for production in the past is irreversible, it is still possible to invest part of the historic emitters' output in capital for future emitters in order to address the concern for distributional equity.

We demonstrate that policy-making faces a fundamental trade-off between Pareto-efficiency and sustainability: one can achieve either one of these two goals, but not both, if policy-making is done initially under unawareness and can be adjusted only after irreversible actions were made. That is, under these conditions one falls short of capturing the maximal potential utility. For climate policy this means that any attempt to achieve climate justice between historic and future emitters necessarily leads to Pareto-inefficiency, and Pareto-efficient policies will not be equitable.

The paper is organized as follows. Section 3.2 introduces the model. In Section 3.3, the normative criteria of sustainability and Pareto-efficiency are defined. Section 3.4 examines the effects of temporal irreversibility and unawareness on policy-making. Section 3.5 discusses the generality and robustness of these results. Section 3.6 concludes and discusses the question of what criteria could provide orientation under irreversibility and unawareness.

3.2 Model

There are two successive, non-overlapping generations $t = 1, 2$. Both have identical preferences over consumption C_t represented by a monotonic and concave utility function $U_t = U(C_t)$. Generation 1 is endowed with stocks of circulating capital and a non-renewable natural resource, with both stocks normalized to 1. Both generations use amounts K_t and R_t of capital and resource for the production of some intermediate good, $Y_t = F(K_t, R_t)$, where F is twice continuously differentiable, concave and exhibits positive and decreasing marginal products of both capital and resource input, $F_{KR} = F_{RK} > 0$, and capital is essential for production, $F(0, R_t) = 0$. Of course, in the absence of any regulation generation 1 will use its capital stock completely $K_1 = 1$. The intermediate good thus produced in $t = 1$ can either be directly consumed by generation 1, or it may be transferred to generation 2 as circulating capital K_2:

$$C_1 = F(1, R_1) - K_2 \tag{1}$$

Generation 2 will use all it inherits from generation 1 in production, K_2 and $R_2 = 1 - R_1$, and it will consume the entire amount of the intermediate good produced in $t = 2$.

The first generation's use of the resource in production causes damages $D(R_1)$ to the second generation, i.e. it diminishes the availability of their social product for consumption,

$$C_2 = (1 - D(R_1))F(K_2, 1 - R_1) \ , \tag{2}$$

with marginal damages being positive and increasing, $D'(R_1) > 0$ and $D''(R_1) \geq 0$, and total damages in the range $0 < D(R_1) < 1$ for all $R_1 > 0$ and $D(0) = 0$. This modeling of damages is close to the DICE model where climate change damages are modeled as a fraction of GDP (Nordhaus 1992). To account for uncertainty on this actual fact, let $\kappa \in \{0, 1\}$ denote the state of information on damages. Second-generation consumption, contingent upon (un)awareness, then is:

$$C_2 = (1 - \kappa D(R_1))F(K_2, 1 - R_1) \ . \tag{3}$$

Initially, i.e. before actual production, generation 1 is unaware of any potential future damages, $\kappa = 0$, and is not even aware of its ignorance, but firmly believes that its resource use does not entail any future damages. That is, they are in a state of "ignorance" (sensu Faber et al. 1992). Thus, future damages are what has been called "unforeseen contingencies" (Dekel et al. 1998) or "unknown unknowns" (Rumsfeld 2002). Only after production by generation 1 has taken place, this unawareness is resolved and the full extent of damages becomes apparent, $\kappa = 1$.

A social planner aims at (1) Pareto-efficiency across generations and (2) sustainability, i.e. non-decreasing utility over time. She acts during the first generation's lifetime and shares the same information as the first generation. In order to achieve her two goals, the social planner has two policy instruments at hand: (1) she can restrict resource use of generation 1, R_1, by an upper limit r; (2) she can oblige generation 1 to transfer at least k out of its intermediate product, Y_1, to generation 2 as capital, K_2.

The exact time structure is as follows. There are three time stages: $t = 1a$, $t = 1b$ and $t = 2$. Generation 1 lives in the first two of these, generation 2 lives in the last one. In the first stage $t = 1a$, generation 1 chooses its capital input K_1, resource input R_1, and capital transfer K_2 so as to maximize its own consumption C_1 subject to restrictions imposed by technology and policy. At this stage, the social planner may restrict resource use by r and make generation 1 plan with a minimal capital transfer of k. Production takes place in this stage, so that the inputs are irreversibly sunk, but as production takes time, the output is not turned out before the next stage. In the second stage $t = 1b$, output Y_1 of the intermediate good becomes available for use. At the same time, uncertainty is resolved and the future damages $D(R_1)$ from using the resource in production become fully apparent. In reaction to this information, the social planner can adjust her policy at this stage. As production by generation 1 has already taken place and resources R_1 are irreversibly sunk, she cannot revise the restriction on resource use any more. However, she can still adjust her second policy instrument and force generation 1 to transfer a higher amount k out of its intermediate good, thereby reducing generation 1's consumption C_1 and increasing generation 2's consumption C_2. Generation 1 cannot revise its production decision anymore at this stage, as the inputs are irreversibly sunk. In the third stage $t = 2$, generation 2 derives utility from its production of the intermediate good Y_2, which is entirely consumed in this same stage, with a reduction due to the damages caused by generation 1's resource use.

Our model is simple, yet captures the key characteristics required to discuss unawareness and irreversibility. In Section 3.5 below, we discuss a number of variations and extensions of this model, to demonstrate generality and robustness of our results.

3.3 Definitions

First, we need to distinguish three definitions of feasibility as there are irreversibility, unawareness, and at a later stage awareness about future damages in the model. Ex-ante (ex-post) feasibility refers to those allocations that are deemed feasible at $t = 1a$ under unawareness (awareness) of future damages. Reduced feasibility refers to those allocations that are feasible at $t = 1b$, i.e. under awareness, after one has acted irreversibly at $t = 1a$. In the following we denote an allocation by $X = (K_1, R_1, Y_1, C_1, K_2, R_2, Y_2, C_2)$.

Definition 1 (Feasibility)

An allocation X is called *ex-ante (ex-post) feasible* if

$$\begin{aligned} 0 \leq K_1 \leq 1,\ 0 \leq K_2 \leq F(K_1, R_1),\ R_1 + R_2 \leq 1,\ R_1, R_2 \geq 0, \\ C_1 = Y_1 - K_2,\ Y_1 = F(K_1, R_1), \\ C_2 = Y_2 = (1 - \kappa D(R_1))F(K_2, 1 - R_1) \ \text{ with } \kappa = 0 \ (\kappa = 1). \end{aligned} \tag{4}$$

For any $0 \leq \overline{K}_1 \leq 1, 0 \leq \overline{R}_1 \leq 1$, and thus $\overline{Y_1} = F(\overline{K_1}, \overline{R_1})$, realized at $t = 1a$, an allocation is called *reduced feasible* if

$$\begin{aligned} 0 \leq K_2 \leq F(\overline{K_1}, \overline{R_1})\ , R_2 \leq 1 - \overline{R_1}\ , R_2 \geq 0\ , \\ C_1 = \overline{Y_1} - K_2, \\ C_2 = Y_2 = (1 - D(\overline{R_1}))F(K_2, 1 - \overline{R_1}). \end{aligned} \tag{5}$$

We understand the terms "sustainability" and "efficiency" as follows. Sustainability is defined as equal utility over time – the minimum requirement for the usual notion of sustainability as non-decreasing utility over time (Howarth 1995). With appropriate specification of the state of information, the criterion is as follows.

Definition 2 (Sustainability)
An ex-ante (ex-post) feasible allocation X is called *ex-ante (ex-post) sustainable* if and only if it is yields

$$U_2 = U_1 \tag{6}$$

$$\text{where } U_1 = U(F(K_1, R_1) - K_2) \tag{7}$$

$$\text{and } U_2 = U((1 - \kappa D(R_1))F(K_2, R_2)) \text{ with } \kappa = 0 \ (\kappa = 1)\ . \tag{8}$$

Similarly, efficiency is defined in an information-and-irreversibility-differentiated manner in the sense of Pareto-efficiency. *Ex-ante efficiency* means that one cannot make a generation better off without making the other worse-off under unawareness of the damages from resource use before any irreversibility in resource use has taken effect. This is the relevant efficiency criterion to guide policy-making in $t = 1a$. *Ex-post efficiency* refers to the hypothetical case where there is awareness of the damages initially, i.e. before any irreversibility has taken effect, so that policy can be fully adjusted to future damages. Thus, it indicates the maximal potential utility in the system that is obtainable under awareness of the inevitable damages of resource use. While initial awareness of damages is hypothetical, it is nevertheless feasible to actually attain ex-post efficient allocations even under initial unawareness - which we will show below.

As a consequence of irreversible resource use in $t = 1a$, there is a reduced set of feasible actions in $t = 1b$. This irreversibility has to be taken into account by the policy-relevant efficiency criterion at this stage: *reduced-feasibility efficiency* encompasses irreversibility and awareness of damages.

Definition 3 (Efficiency)
a) An ex-ante (ex-post) feasible allocation X is called *ex-ante (ex-post) efficient* if and only if there exists no other ex-ante (ex-post) feasible allocation X' for which $U_t' \geq U_t$ for $t = 1, 2$ and $U_t' > U_t$ for at least one t.
b) Contingent on $K_1 = \overline{K_1}, R_1 = \overline{R_1}, Y_1 = \overline{Y_1}$, a reduced feasible allocation X is called *reduced-feasibility efficient* (for short: *RF-efficient*) if and only if there exists no other reduced-feasible allocation X' with $K_1' = \overline{K_1}, R_1' = \overline{R_1}, Y_1' = \overline{Y_1}$ for which $U_t' \geq U_t$ for $t = 1, 2$ and $U_t' > U_t$ for at least one t.

With this definition, efficient allocations are characterized as follows.

Lemma 1
An ex-ante (ex-post, reduced) feasible allocation X is ex-ante (ex-post, reduced) efficient if and only if it meets the following conditions:
(i) Ex-ante efficiency:

$$\begin{aligned} F_R(1, R_1) F_K(K_2, 1 - R_1) &= F_R(K_2, 1 - R_1), && (9) \\ F(K_2, 1 - R_1) &= \overline{C} \ \text{ with } \overline{C} \in [0, \overline{C}^{EA,max}], && (10) \end{aligned}$$

(ii) Ex-post efficiency:

$$F_R(1, R_1)F_K(K_2, 1 - R_1) - D'(R_1)F(K_2, 1 - R_1)/(1 - D(R_1)) = F_R(K_2, 1 - R_1), \tag{11}$$

$$(1 - D(R_1))F(K_2, 1 - R_1) = \overline{C} \ \text{ with } \overline{C} \in [0, \overline{C}^{EP,max}], \tag{12}$$

(iii) Reduced-feasibility efficiency:

$$(1 - D(\overline{R_1}))F(K_2, 1 - \overline{R_1}) \ = \ \overline{C} \ \text{ with } \overline{C} \in [0, \overline{C}^{RF,max}], \tag{13}$$

where $\overline{C}$ is an intergenerational distribution parameter with $\overline{C} \in [0, \overline{C}^{i,max}]$, $i \in \{EA, EP, RF\}$.

Proof. See Appendix A.1. □

As consumption is the only good in the model, the intergenerational distribution parameter $\overline{C}$ determines not only consumption, but also utility levels. It can attain any value between 0 (all potential utility in this system is with generation 1, none with generation 2) and some $\overline{C}^{i,max}$ (all potential utility is with generation 2, none with generation 1), so that there exist infinitely many efficient allocations satisfying these characterizations. Conditions (9), and (11) state that the marginal gain in consumption for generation 2 from either of the following two alternative uses of the resource should be equal: (LHS) giving the additional resource to generation 1 as input into production, and then transferring the entire additional amount of the intermediate good thus produced as capital into generation 2's production; (RHS) giving the additional resource directly to generation 2 as input into their production. While Condition (9) expresses this without taking damages into account ($\kappa = 0$), Condition (11) states the same for the case where the damages are known ($\kappa = 1$) and taken into account from the beginning.[2] Condition (13) states that varying $\overline{C}$ determines the RF-efficient capital transfer K_2 which generates the set of RF-efficient allocations.

One can illustrate the efficient allocations through continuous and monotonically decreasing utility frontiers $U_2(U_1)$ in utility space (Figure 3.7). Unawareness (awareness)

[2] As (11) differs from (9), ex-ante efficient allocations are, in general, ex-post inefficient.

at $t = 1a$ about the damages yields the ex-ante (ex-post) utility frontier which runs from $U_2^{EA,max}$ ($U_2^{EP,max}$) to U_1^{max}.

Lemma 2
The ex-post utility frontier is the envelope of the reduced-feasibility utility frontiers that result for all $\overline{R_1} \in [0, 1]$.

Proof. See Appendix A.2. □

The ex-post utility frontier forms the envelope of all RF-utility frontiers. The RF-utility frontier depends on the actual realization of $R_1 = \overline{R_1}$ at $t = 1a$ and runs from $U_2^{RF,max}$ to $U_1^{RF,max}$. Due to the envelope property of the ex-post frontier and unawareness about factual damages, the ex-ante and ex-post frontiers both Pareto-dominate the RF-utility frontier. In the same figure, sustainable allocations are represented as a 45°-line from the axes.

3.4 Results

Under a laissez-faire policy the first generation will use up its entire circulating capital $K_1^0 = 1$, exploit the total amount of the resource $R_1^0 = 1$, and will not provide a capital transfer to the next generation, $K_2^0 = 0$. As a consequence, the first generation is better-off than the second one $U_1^0 > U_2^0$, as $C_1^0 = F(1, 1) > 0$ and $C_2^0 = F(0, 0) = 0$ (illustrated by X^0 in Figure 3.7). While this laissez-faire allocation is efficient by any notion of efficiency, it is not sustainable. This motivates sustainability policy by the social planner.

Sustainability policy has to follow the time structure laid out in Section 3.2. At $t = 1a$, the social planner devises a policy mix of restrictions on resource use and a capital transfer which should lead to an ex-ante efficient and ex-ante sustainable allocation.

Proposition 1 (Ex-ante sustainable and ex-ante efficient policy)
At time $t = 1a$, there exists a unique policy mix $(\tilde{r}, \tilde{k})$ that leads to an allocation

$\tilde{X}$ which is both ex-ante sustainable and ex-ante efficient. It is characterized by the following necessary and sufficient conditions:

$$\begin{aligned} F_R(1,\tilde{r})F_K(\tilde{k},1-\tilde{r}) &= F_R(\tilde{k},1-\tilde{r})\ , && (14)\\ F(1,\tilde{r})-\tilde{k} &= F\left(\tilde{k},1-\tilde{r}\right)\ . && (15)\end{aligned}$$

Proof. See Appendix A.3. □

In climate policy, the policy mix $(\tilde{r},\tilde{k})$ refers to the situation under which production and consumption of the historic emitters took place. The use of fossil fuels was thought to be harmless as the effects of greenhouse gases on the atmosphere were unknown unknowns. Future emitters were thought to be at least equally well-off due to capital transfers and the ability to use the remaining fossil fuels. Yet, with the first report of the IPCC (1990) there was strong evidence that anthropogenic climate change was real and of relevant magnitude, and estimates of the damages due to historic greenhouse-gas emissions became available.

Thus, in the second stage $t = 1b$ the damages from resource use $D(R_1)$ are known, $\kappa = 1$. Obviously, not adapting the policy mix $(\tilde{r},\tilde{k})$ to the new findings would result in an ex-post unsustainable allocation, with generation 2 worse-off than generation 1, as $U(F(1,\tilde{r})-\tilde{k}) > U((1-D(\tilde{r}))F(\tilde{k},1-\tilde{r}))$. Therefore, the social planner must adapt her policy mix to ensure ex-post sustainability. However, production of the first generation of $\tilde{Y}_1 = F(\tilde{K}_1,\tilde{R}_1)$ has already taken place and inputs are irreversibly sunk. The only viable instrument is, therefore, to increase the minimum transfer of capital from generation 1 to generation 2, $k > \tilde{k}$. As this does not allow the achievement of ex-post efficiency, the policy maker faces the following fundamental sustainability-efficiency trade-off.

Proposition 2 (Sustainability-efficiency trade-off)
In general, and in particular for the ex-ante efficient and ex-ante sustainable policy $(\tilde{r},\tilde{k})$, policy-making at time $t = 1b$ faces the following trade-off between ex-post efficiency and ex-post sustainability:

(i) there exists no policy mix $(r = \tilde{r}, k)$ that yields an allocation that is both ex-post efficient and ex-post sustainable, but

(ii) there exists a unique policy mix $(\hat{r} = \tilde{r}, \hat{k})$ with $\hat{k} > \tilde{k}$ that yields an allocation $\hat{X}$ that is ex-post sustainable but not ex-post efficient, and

(iii) there exists another unique policy mix $(r^* = \tilde{r}, k^*)$ with $k^* < \tilde{k}$ that yields an allocation X^* that is ex-post efficient but not ex-post sustainable.

Proof. See Appendix A.4. □

The intuition behind this result is as follows. Despite the damages the social planner can still achieve an ex-post sustainable allocation at time $t = 1b$ by adjusting her policy mix (Proposition 2 ii). This requires generation 1 to transfer more of its intermediate product as circulating capital to the second generation than originally planned, $\hat{k} > \tilde{k}$. As this transfer would exceed the one originally deemed necessary for sustainability and generation 1's production is irreversible, Condition (11) for ex-post efficiency is violated. Still, sustainability is achieved in spite of the damages and the irreversibility of resource use, as $U(\hat{C_1}) = U(F(1, \tilde{r}) - \hat{k}) = U((1 - D(\tilde{r}))F(\hat{k}, 1 - \tilde{r})) = U(\hat{C_2})$. In climate policy this corresponds to a higher transfer of capital from historic to future emitters to compensate them for (previously unknown) damages from climate change.

Alternatively, an ex-post efficient allocation X^* can be achieved, by decreasing the capital transfer in $t = 1b$, i.e. $k^* < \tilde{k}$ and $r^* = \tilde{r}$ (Proposition 2 iii). It is, however, not ex-post sustainable as $U(C_1^*) = U(F(1, r^*) - k^*) > U((1 - D(r^*))F(k^*, 1 - r^*)) = U(C_2^*)$. For climate policy this would mean a lower capital transfer and therefore no compensation to future emitters for damages from climate change. As the minimum capital transfer k is the only remaining policy variable at time $t = 1b$, and $k = \hat{k}$ would ensure ex-post sustainability while any $k \neq \tilde{k}$ leads to ex-post inefficiency, there exists no k that achieves both ex-post sustainability and ex-post efficiency (Proposition 2 i).

Despite this fundamental trade-off in policy-making with the two policy instruments studied here – a limit r on resource use by generation 1 and a minimum capital transfer k from generation 1 to generation 2 – there exists, in principle, a reduced feasible allocation that is both ex-post efficient and ex-post sustainable.

Proposition 3 (Bliss)
There exists a unique policy mix (r^{Bliss}, k^{Bliss}) that yields an ex-post efficient and ex-post sustainable allocation X^{Bliss} which is reduced feasible, that is, feasible under unawareness and irreversibility.

Proof. See Appendix A.5. □

This result mirrors the one from the intragenerational equity-efficiency trade-off where a first-best efficient and equitable allocation is feasible under all physical constraints, but not achievable with given instruments of social policy.

The various policies are illustrated with regard to sustainability and efficiency in Figure 3.7. The laissez-faire allocation X^0 without policy interference is unsustainable, but ex-ante and ex-post efficient. The ex-ante efficient and ex-ante sustainable allocation $\tilde{X}$ that results from policy $(\tilde{r}, \tilde{k})$ in $t = 1a$ lies at the intersection of the ex-ante utility frontier (dashed curve) with the sustainability line (45°-line). When damages to generation 2 become apparent in $t = 1b$ and the capital transfer is reduced to k^*, this allocation becomes the ex-post efficient, yet ex-post unsustainable allocation X^*.

All possible redistributions through a capital transfer k from generation 1 to generation 2 at $t = 1b$ after irreversibility in resource use $\tilde{r}$ has taken effect, generate the RF-utility frontier (solid curve). Depending on whether the transfer is increased $k > \tilde{k}$ or decreased $k < \tilde{k}$ one moves along the RF-utility frontier closer or further away from the sustainability line. The RF-utility frontier lies strictly below the ex-post utility frontier (dotted curve) except at X^* where both coincide (for $k = k^*$). This RF-utility frontier allows attaining the ex-post sustainable, yet ex-post inefficient allocation $\hat{X}$. Beyond the RF-utility frontier, there exists the ex-post sustainable and ex-post efficient allocation X^{Bliss} at the intersection of the ex-post utility frontier and the sustainability line.

The trade-off between ex-post sustainability and ex-post efficiency at stage $t = 1b$ consists in the impossibility to reach both goals at once, i.e. the social planner must choose between the ex-post efficient allocation X^* and the ex-post sustainable allocation $\hat{X}$. She can also choose any combination of the two on the RF-utility frontier.

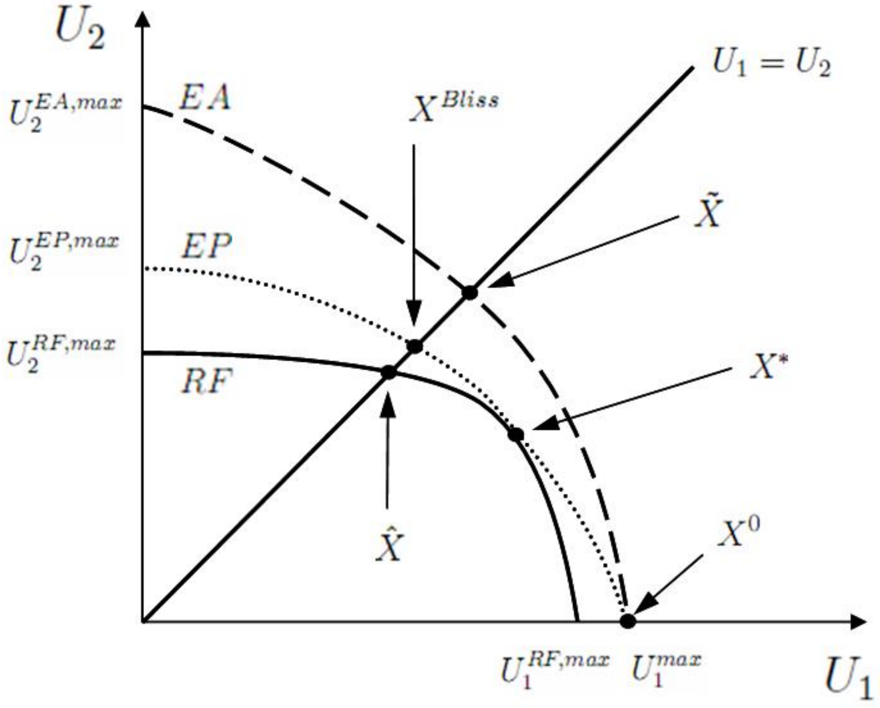

Figure 3.7: Illustration in the space of future utility U_2 and present utility U_1 of sustainable allocations (45°-line); ex-ante efficient allocations (EA, dashed); ex-post efficient allocations (EP, dotted), including the ex-post efficient and ex-post sustainable allocation X^{Bliss}; and reduced-feasibility efficient allocations (RF, solid), highlighting the trade-off between an ex-post sustainable allocation $\hat{X}$ and an ex-post efficient allocation X^*.

Without irreversibility or unawareness there would be no such trade-off: if there were no irreversibility the ex-post utility frontier would be attainable in $t = 1b$ as the policy mix could easily be adjusted to the previously unknown damages. If there was no unawareness of future damages the social planner would simply choose a sustainable allocation on the ex-post utility frontier in $t = 1a$. Here, irreversibility would not be an issue as sustainability problems would be apparent from the very beginning.

On a more general note, Figure 3.7 illustrates the welfare loss due to the combination of irreversibility and unawareness. This is represented as the difference between the EP-utility frontier that indicates the maximal potential utility and the RF-utility frontier that indicates achievable allocations after irreversible actions were taken. It becomes clear that every irreversible action (such as e.g. the depletion of natural resource stocks, or the generation of persistent pollutants and wastes) implies a welfare loss if an initially unknown negative effect becomes apparent at a later stage. This is due to the fact that correcting an irreversible decision afterwards is done with even more limited means as some resources have irreversibly been used.

3.5 Discussion

In this section, we discuss a number of common variations and extensions of the model (cf. Section 3.2) to show the generality and robustness of our results.

Positive externality

In the model there is unawareness about a negative intergenerational externality that originates from resource use. If there is unawareness about a positive externality that benefits the second generation, e.g. climate change might be beneficial to (at least some parts of) future generations (cf. Tol 2009), policy runs into the same trade-off: with initial use of the resource being too low, there is still a trade-off between ex-post efficiency and ex-post sustainability after irreversibility has taken effect and unawareness is resolved. In case there are positive and negative externalities from resource use the net of the two matters – and as long they do not cancel out the same result obtains.

Sustainability as non-decreasing utility

For reasons of uniqueness we chose a definition of sustainability as equal utility between generations. If sustainability is defined more generally as non-decreasing utility over time (as e.g. by Solow 1974, Hartwick 1977, Solow 1986, Howarth 1995, Arrow et al. 2004), the equity-efficiency trade-off persists up to a boundary case. If resource use in the first period is such that the intersection between the resulting RF-utility frontier and the EP-utility frontier coincides with the allocation X^{Bliss}, there is no trade-off between ex-post efficiency and ex-post sustainability. Any resource use by the first generation below this level also leads to RF-utility frontiers which include an ex-post sustainable and ex-post efficient allocation. Yet, under unawareness there is no information so as to characterize this boundary case and achieving sustainability and efficiency simultaneously would be mere chance.

Infinite number of generations

In the model there are two successive generations, but our results hold for an infinite number of generations (as assumed by e.g. Svenson 1980, Solow 1986, Asheim and Tungodden 2004) as long as at least one of them is subject to unawareness and irreversibility. If this generation takes an irreversible decision under unawareness, policy-making – irrespective of the number of generations – subsequently faces the trade-off between ex-post efficiency and ex-post sustainability.

Overlapping generations

In the model generations do not overlap. An overlapping-generations model would split the lifetime of each generation in a young production stage and an old consumption stage with an overlap between the two generations (e.g. Howarth and Norgaard 1992, Burton 1993, Marini and Scaramozzino 1995, Howarth 1998). If the first generation takes irreversible production decisions under unawareness in its young production stage, it cannot change its past production in its consumption stage – even if it overlaps with the successive generation. In fact this is quite close to the model where the lifetime of generation 1 is split in two periods (1a, 1b). This means that as long as irreversibility and unawareness persist in the production decision of the first generation, policy-making faces an equity-efficiency trade-off with overlapping generations.

Amenity value and other services of the resource

If the first generation directly draws utility from the resource stock, e.g. due to some amenity value (e.g. Krautkraemer 1985), it will of course reduce its resource use. Yet, if there is unawareness about the intergenerational externality, policy cannot efficiently adapt and overuse of the resource remains. Other services of the resource stock affect resource use of the first generation similarly, but this cannot internalize the unknown damages from resource use. Here, the degree of substitutability between amenity values and consumption does not matter (as in the discussion of weak vs. strong sustainability, e.g. Neumayer 1999) as long as the resource is used as a production input at least to some extent.

Non-resource-based goods

The model features one single aggregate consumption good which is produced from the resource. In particular, there is no non-resource-based good that could serve as a substitute for consumption, such as e.g. leisure in the leisure-consumption model (e.g. Sandmo 1981). If a generation optimizes intragenerationally with respect to several goods, the combination of irreversibility and unawareness in one of those goods violates first-best efficiency (leading to an equity-efficiency trade-off) just as with resource use in the model presented above.

Renewable resource

The model features a non-renewable resource. This makes sustainability impossible over an infinite time-horizon. If the natural resource is renewable (e.g. Clark 2010), the social planner incorporate this in her ex-ante efficient and ex-ante sustainable policy. If the use or consumption of this resource is irreversible and causes an ex-ante unforeseen effect, there exists an equity-efficiency trade-off.

Fixed capital

If capital is not used up in production process, but can be accumulated (as e.g. in the model due to Dasgupta and Heal 1974, Solow 1974, Stiglitz 1974) unawareness still poses a problem for efficiency. For example, an efficient consumption path with a respective built-up in capital cannot be adjusted to an intergenerational externality without losing efficiency if it becomes apparent after irreversibility has taken effect.

Technical progress

An important factor in intertemporal problems is technical progress which benefits future generations (e.g. Aghion and Howitt 1998: Ch. 5, Schou 2002, Acemoglu et al. 2012). Technical progress, to the extent that it is known, would be included in the ex-ante efficient and ex-ante sustainable policy by the social planner. With exogenous technical progress this policy redistributes less resource and capital to the second generation ex ante. With endogenous technical progress this policy requires the first generation to invest ex ante in technical progress. But both variants of technical progress do not compensate the future generation for an unknown externality. If there is unawareness of technical progress this leads to the discussion of a positive externality from above.

Source of irreversibility

In the model, the use of the natural resource in production is irreversible. Other sources of irreversibility are imaginable. For instance, in a leisure-consumption model where individuals produce consumption from labor as the sole production factor (e.g. Sandmo 1981), time is irreversible. Thus, leisure cannot be transferred over time but is irreversibly used by each generation. In this model, if the social planner surprises generation by announcing an unforeseen tax or transfer (e.g. Atkinson and Stiglitz 1980: 68) after leisure has irreversibly been enjoyed, there exists an intergenerational equity-efficiency trade-off.

Object of unawareness

In the model there is unawareness of future damages and, thus, of production possibilities. Other objects of potential unawareness include future preferences, resource extraction costs or resource stock size. Obviously, if unawareness of either of these objects is resolved after irreversible production has taken place, the trade-off between ex-post sustainability and ex-post efficiency remains.

Altruism

The role of altruism for long-run sustainability is often discussed (e.g. Jouvet et al. 2000, Bréchet and Lambrecht 2009). In our model the social planner, rather than the generations, pursues the normative criteria of efficiency and sustainability. If the first generation is altruistic towards the second generation, it would by itself provide a transfer to the future. Yet, due to unawareness, this voluntary transfer does not account for the negative externality from resource use. Subsequently, irreversibility in production does not allow the altruistic first generation to readjust which leads to the equity-efficiency trade-off as before.

Kind of transfer

In distributional problems the kind of transfer matters, i.e. whether income, consumption, or capital is redistributed. Our model uses a transfer of capital which is a durable good that can be redistributed across generations. Other kinds of transfers such as income or consumption would lead to the same result: if there is irreversibility in production and unawareness about future damages, none of these transfers can ensure equity and efficiency simultaneously.

This discussion highlights the importance of the combination of unawareness and irreversibility as the key characteristics driving the results. As for all other assumptions of the model, the results are robust to a large number of variations and extensions.

3.6 Conclusion

We have studied the question of whether there exists a mechanism genuine to intergenerational policy-making that causes an intergenerational equity-efficiency trade-off. We found that sustainability policy that acts under a combination of temporal irreversibility and unawareness faces such a trade-off between efficiency across generations and intergenerational equity. Hence, in general it falls short of capturing the maximal potential utility from the system.

This result is relevant for current climate policy. Policies that want to achieve sustainability after damages were initially unknown (unawareness) must respect that past actions cannot be undone (temporal irreversibility), and that redistribution therefore faces a trade-off between efficiency and sustainability. For the case of climate justice – where climate policy is enacted after production and emissions have already irreversibly taken place – this means that there is a trade-off between equity and efficiency among historic and future emitters. Policymakers therefore need to be aware of the fact that pursuing sustainability as the overriding priority sacrifices efficiency, and that prudent policy-making requires a prior debate on how to balance these two conflicting goals.

Our result is very general and holds way beyond the model studied here and beyond the case of climate policy. For, one simply cannot think of intergenerational policy-making that is *not* subject to irreversibility and unawareness. Hence, any intergenerational policy-making is, in general, subject to an intergenerational equity-efficiency trade-off. For instance, this holds also for industrial and technology policy, or the design of pension systems.

This raises the question of how one should act in the face of irreversibility and unawareness. Policy that explicitly aims at the two normative objectives of equity and efficiency at all stages of history cannot attain both of them. Yet, as a "bliss" allocation is actually feasible, even under irreversibility and unawareness, one would expect relevant normative criteria to guide us there. But in the second-best world in which one necessarily acts, one is without orientation with respect to the two normative objectives of equity and efficiency.

Against this background, several conclusions emerge: First, beware of irreversibility. As irreversibility reduces the possibilities of reacting to unforeseen developments – both negative and positive – one would be better off with less irreversibility. And irreversibility may indeed be evident ex-ante and a matter of choice. For example, technologies of electricity production, e.g. from wind or nuclear fuels, clearly differ in terms of irreversibility.

Second, beware of unawareness. As with irreversibility, unawareness may be evident ex-ante and a matter of choice. For instance, unawareness may be reducible or not (Faber et al. 1992). In the latter case, there is nothing one can do to reduce unawareness. In the former case, research may allow the reduction of unawareness – to narrow the gap between ex-post and ex-ante frontiers. As an example, in the beginning of the industrialization, people were unaware about the potential consequences of releasing carbon emissions into the atmosphere, and they were not even aware of their unawareness. Later, in the 1950s, it became evident to some that there was unawareness about the potential consequences of releasing carbon emissions into the atmosphere. Climate research since then has outlined these potential states of the world, thus reducing unawareness.

Third, in instances of unawareness orientation can come from weaker normative objectives. They should build on conceptions and variables of which one is not fundamentally unaware, but of which one can – with good epistemological reasons – believe to have more knowledge. For example, sustainability as considered here – non-declining utility over time – is a very knowledge-demanding criterion, because it requires full knowledge of future preferences, production technology, system dynamics, etc. A weaker criterion of sustainability could be that of equal satisfaction of basic needs (sensu WCED 1987). Knowledge demand for assessing this criterion is considerably lower, and less prone to fundamental unawareness, as one does not need to know individual preferences. Likewise, one could consider weaker and less knowledge-demanding criteria of efficiency than Pareto-efficiency, e.g. non-wastefulness in production and transfer.

Fourth, the fundamental problem that unawareness poses for the evaluation of consequences of one's action, does not exist in deontological ethics, which justifies actions without relying on information about their consequences. Some have suggested such deontological criteria especially with regard to sustainability (e.g. Howarth 1995).

Nevertheless, analyzing the efficiency of instruments in sustainability policy (as e.g. in Gerlagh and Keyzer 2001) is still indispensable. Describing and quantifying the trade-offs between sustainability and efficiency helps to outline the limits for the design of concrete policies. After all, we do not want to pay more for sustainability than necessary – even in the face of irreversibility and unawareness.

Acknowledgements

We are grateful to Geir Asheim and two anonymous referees from a peer-reviewed journal for exceptionally insightful and helpful comments, to Maik Heinemann for critical and constructive discussion, and to the German Federal Ministry of Education and Research for financial support under grant 01UN1011A.

Bibliography

Acemoglu, D., Aghion, A., Bursztyn, L. and Hemous, D. (2012). The environment and directed technical change. *American Economic Review*, 102(1):131-166.

Aghion, P. and Howitt, P.W. (1998). *Endogenous Growth Theory.* MIT Press, Cambridge.

Arrow, K., Dasgupta, P., Goulder, L., Daily, G., Ehrlich, P., Heal, G., Levin, S., Mäler, K.-G., Schneider, S., Starrett, D. and Walker, D. (2004). Are we consuming too much? *Journal of Economic Perspectives*, 18(3):147–172.

Asheim, G.B. and Tungodden, B. (2004). Resolving distributional conflicts between generations. *Economic Theory*, 24(1):221–230.

Atkinson, A.B. and Stiglitz, J.E. (1980). *Lectures on Public Economics.* McGraw-Hill, London.

Baumgärtner, S. (2005). Temporal and thermodynamic irreversibility in production theory. *Economic Theory*, 26(3):725–728.

Baumgärtner, S. and Quaas, M.F. (2010). Sustainability economics – general versus specific, and conceptual versus practical. *Ecological Economics*, 69(11):2056–2059.

Bréchet, T. and Lambrecht, S. (2009). Family altruism with renewable resource and population growth. *Mathematical Population Studies*, 16(1):60–78.

Burton, P.S. (1993). Intertemporal preferences and intergenerational equity considerations in optimal resource harvesting. *Journal of Environmental Economics and Management*, 24(2):119–132.

Clark, C.W. (2010). *Mathematical Bioeconomics. The Mathematics of Conservation.* 3rd edition, Wiley, New York.

Dasgupta, P. and Heal, G. (1974). The optimal depletion of exhaustible resources. *Review of Economic Studies*, 41:3–28.

Dekel, E., Lipman, L. and Rustichini, A. (1998). Recent development in modeling unforeseen contingencies. *European Economic Review*, 42(3):523–542.

Faber, M., Manstetten, R. and Proops, J.L.R. (1992). Humankind and the environment: An anatomy of surprise and ignorance. *Environmental Values*, 1(3):217–242.

Gerlagh, R. and Keyzer, A.M. (2001). Sustainability and the intergenerational distribution of natural resource entitlements. *Journal of Public Economics*, 79(2):315–341.

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

Howarth, R.B. (1995). Sustainability under uncertainty: a deontological approach. *Land Economics*, 71(4):417–427.

Howarth, R.B. (1998). An overlapping generations model of climate-economy interactions. *Scandinavian Journal of Economics*, 100(3):575–591.

Howarth, R.B. and Norgaard, R.B. (1990). Intergenerational resource rights, efficiency, and social optimality. *Land Economics*, 66(1):1–11.

Howarth, R.B. and Norgaard, R.B. (1992). Environmental valuation under sustainable development. *American Economic Review – Papers and Proceedings*, 82(2):473–477.

Intergovernmental Panel on Climate Change (IPCC) (1990). *First Assessment Report.* Cambridge University Press, Cambridge.

Intergovernmental Panel on Climate Change (IPCC) (2007). *Climate Change 2007: The Physical Science Basis; Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge.

Jouvet, P.-A., Michel, P. and Vidal, J.-P. (2000). Intergenerational altruism and the environment. *Scandinavian Journal of Economics*, 102(1):135–150.

Krautkraemer, A.J. (1985). Optimal growth, resource amenities and the preservation of natural environments. *Review of Economic Studies*, 52(1):153–170.

Krautkraemer, A.J. and Batina, G.R. (1999). On sustainability and intergenerational transfers with a renewable resource. *Land Economics*, 75(2):167–184.

Krysiak, F.C. (2009). Sustainability and its relation to efficiency under uncertainty. *Economic Theory*, 41(2):297–315.

Le Grand, J. (1990). Equity versus efficiency: the elusive trade-off. *Ethics*, 100(3):554-568.

Marini, G. and Scaramozzino, P. (1995). Overlapping generations and environmental control. *Journal of Environmental Economics and Management*, 29(1):64–77.

Millenium Ecosystem Assessment (MEA) (2005). *Ecosystems and Human Well-Being.* Island Press, Washington, DC.

Neumayer, E. (1999). *Weak Versus Strong Sustainability: Exploring the Limits of Two Opposing Paradigms.* Edward Elgar, Cheltenham.

Neumayer, E. (2000). In defence of historical accountability for greenhouse gas emissions. *Ecological Economics*, 33(2):185–192.

Nordhaus, W. (1992). An optimal transition path for controlling greenhouse gases. *Science*, 285(5086):1315-1319.

Pezzey, J.C.V. (2004). Sustainability policy and environmental policy. *Scandinavian Journal of Economics*, 106(2):339–359.

Putterman, L., Roemer, J.E. and Silvestre, J. (1998). Does egalitarianism have a future? *Journal of Economic Literature*, 36(2):861–902.

Rumsfeld, D. (2002). US Department of Defense News Briefing on February 12, 2002. Available at http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636.

Sandmo, A. (1981). Income tax evasion, labour supply, and the equity-efficiency trade-off. *Journal of Public Economics*, 16(3):265–288.

Schou, P. (2002). When environmental policy is superfluous: Growth and polluting resources. *Scandinavian Journal of Economics*, 104(4):605–620.

Svenson, L.-G. (1980). Equity among generations. *Econometrica*, 48(5):1251–1256.

Solow, M.R. (1974). Intergenerational equity and exhaustible resources. *Review of Economic Studies*, 41:29–45.

Solow, M.R. (1986). On the Intergenerational allocation of natural resources. *Scandinavian Journal of Economics*, 88(1):141–149.

Stiglitz, J. (1974). Growth with exhaustible natural resources: Efficient and optimal growth paths. *Review of Economic Studies*, 41:123–137.

Tol, R.S.J. (2009). The economic effects of climate change. *Journal of Economic Perspectives*, 23(2):29–51.

United Nations Environment Programme (UNEP) (2012). *Global Environment Outlook 5: Environment for the Future We Want.* UNEP, Nairobi, Kenya.

World Commission on Environment and Development (WCED) (1987). *Our Common Future.* Oxford University Press, Oxford.

Appendix

A.1 Proof of Lemma 1

As consumption is the only good in the model, Pareto-efficiency can be analyzed on the level of consumption. Thus, the intergenerational distribution parameter $\overline{C}$ equally determines the utility levels.

(i) An ex-ante feasible ex-ante efficient allocation is the solution to

$$\max_{K_1,R_1,K_2,R_2} C_1 \text{ s.t. } C_2 \geq \overline{C},\ C_1 = F(K_1,R_1) - K_2,\ C_2 = F(K_2,R_2),\ R_1 + R_2 = 1,\ K_1 = 1, \tag{A.16}$$

with the Lagrangian $\mathfrak{L} = F(1,R_1) - K_2 + \lambda_1(F(K_2, 1-R_1) - \overline{C})$. Obviously, in the optimal solution the constraint $C_2 \geq \overline{C}$ must hold with equality. The necessary first order conditions then are:

$$\frac{\partial \mathfrak{L}}{\partial R_1} = F_R(1,R_1) + \lambda_1 F_R(K_2, 1-R_1)(-1) = 0\ , \tag{A.17}$$

$$\frac{\partial \mathfrak{L}}{\partial K_2} = (-1) + \lambda_1 F_K(K_2, 1-R_1) = 0\ , \tag{A.18}$$

$$\frac{\partial \mathfrak{L}}{\partial \lambda_1} = F(K_2, 1-R_1) - \overline{C} = 0\ . \tag{A.19}$$

Rearranging (A.17) and (A.18) and eliminating λ_1 by dividing the two, one arrives at (9). (A.19) yields (10). These conditions are also sufficient, as the optimization problem (A.16) is strictly convex.

(ii) An ex-post feasible ex-post efficient allocation is the solution to

$$\max_{K_1,R_1,K_2,R_2} C_1 \text{ s.t. } C_2 \geq \overline{C},\ C_1 = F(K_1,R_1) - K_2,$$

$$C_2 = (1 - D(R_1))F(K_2,R_2),\ R_1 + R_2 = 1,\ K_1 = 1, \tag{A.20}$$

with the Lagrangian $\mathfrak{L} = F(1,R_1) - K_2 + \lambda_2((1-D(R_1))F(K_2,1-R_1) - \overline{C})$. Obviously, in the optimal solution the constraint $C_2 \geq \overline{C}$ must hold with equality. The necessary first order conditions then are:

$$\frac{\partial \mathfrak{L}}{\partial R_1} = F_R(1,R_1) + \lambda_2(-D'(R_1)F(K_2,1-R_1) + (1-D(R_1))F(K_2,1-R_1)(-1)) \tag{A.21}$$

$$\frac{\partial \mathfrak{L}}{\partial K_2} = (-1) + \lambda_2(1-D(R_1))F_K(K_2,1-R_1) = 0\ , \tag{A.22}$$

$$\frac{\partial \mathfrak{L}}{\partial \lambda_1} = (1-D(R_1))F(K_2,1-R_1) - \overline{C} = 0\ . \tag{A.23}$$

Rearranging (A.21) and (A.22) and eliminating λ_1 by dividing the two, one arrives at (11). (A.23) yields (12). These conditions are also sufficient, as the optimization problem (A.20) is strictly convex.

(iii) A reduced feasible RF-efficient allocation with given $K_1 = \overline{K_1}, R_1 = \overline{R_1}$ is the solution to

$$\max_{K_2} C_1 \text{ s.t. } C_2 \geq \overline{C},\ C_1 = F(\overline{K_1}, \overline{R_1}) - K_2,\ C_2 = (1 - D(\overline{R_1}))F(K_2, \overline{R_2}), \tag{A.24}$$
$$\overline{R_1} + \overline{R_2} = 1,\ \overline{K_1} = 1, 0 \leq K_2 \leq K_2^{RF,max} = F(\overline{K_1}, \overline{R_1})$$

with the Lagrangian $\mathfrak{L} = F(1, \overline{R_1}) - K_2 + \lambda((1 - D(\overline{R_1}))F(K_2, 1 - \overline{R_1}) - \overline{C})$. Obviously, in the optimal solution the constraint $C_2 \geq \overline{C}$ must hold with equality. The necessary first order conditions then are:

$$\frac{\partial \mathfrak{L}}{\partial K_2} = -1 + \lambda(1 - D(\overline{R_1}))F_K(K_2, 1 - \overline{R_1}) = 0\ , \tag{A.25}$$
$$\frac{\partial \mathfrak{L}}{\partial \lambda} = (1 - D(\overline{R_1}))F(K_2, 1 - \overline{R_1}) - \overline{C} = 0\ , \tag{A.26}$$

Choosing $\overline{C}$ determines K_2 in (A.26). K_2 determines λ in (A.25). Thus, (A.26) yields (13). Varying $\overline{C}$ between $\overline{C}^{min} = 0$ and $\overline{C}^{RF,max}$ by varying K_2 between 0 and $K_2^{max} = F(\overline{K_1}, \overline{R_1})$ generates the set of RF-efficient allocations.

A.2 Proof of Lemma 2

Decision problem (A.20) is the envelope to decision problem (A.24). Note, that this analysis holds equally for utility as consumption is the only good in the model. To prove this, define the value function $v(K_2)$ as the solution to the optimization problem (A.20):

$$v(K_2) = F(1, R_1(K_2)) - K_2. \tag{A.27}$$

The derivative of $v(K_2)$ is:

$$\frac{\partial v(K_2)}{\partial K_2} = F_R(1, R_1)\frac{\partial R_1}{\partial K_2} + (-1) \tag{A.28}$$

From (A.20) define constraint function $g(R_1(K_2), K_2) = (1 - D(R_1(K_2)))F(K_2, 1 - R_1(K_2)) - \overline{C}$ which is 0 for all K_2.

From (A.21) we know:

$$F_R(1, R_1) = \lambda_2 \frac{\partial g(R_1(K_2), K_2)}{\partial R_1} \tag{A.29}$$

Inserting this into (A.28) leads to

$$\frac{\partial v(K_2)}{\partial K_2} = \lambda_2 \frac{\partial g(R_1(K_2), K_2)}{\partial R_1} \frac{\partial R_1}{\partial K_2} + (-1) \tag{A.30}$$

As $dg(R_1(K_2), K_2)/dK_2 = 0$ this leads to:

$$\frac{\partial g(R_1(K_2), K_2)}{\partial R_1} \frac{\partial R_1}{\partial K_2} = -\frac{\partial g(R_1(K_2), K_2)}{\partial K_2} \tag{A.31}$$

Inserting this into (A.30) this leads to:

$$\frac{\partial v(K_2)}{\partial K_2} = -1 - \lambda_2 \frac{\partial g(R_1(K_2), K_2)}{\partial K_2} = -1 - \lambda_2((1 - D(R_1))F_K(K_2, 1 - R_1)) \tag{A.32}$$

For Pareto-efficiency the value function must be maximized which requires $\partial v(R_1^*, K_2)/\partial K_2 = 0$ and results the same condition as in (A.25). Therefore, the ex-post utility frontier forms the envelope of the RF-utility frontier.

A.3 Proof of Proposition 1

At time $t = 1a$, the social planner sets (r, k), expecting – given her unawareness, $\kappa = 0$ – that both generations, when maximizing their individual consumption subject to constraints from technology and policy, end up in an allocation $X = (K_1, R_1, Y_1, C_1, K_2, R_2, Y_2, C_2)$ with

$$\begin{aligned} C_1 &= F(1, r) - k & \text{(A.33)} \\ C_2 &= F(k, 1 - r)\ . & \text{(A.34)} \end{aligned}$$

Note, that this analysis holds equally for utility as consumption is the only good in the model. The social planner chooses (r, k) so that X is ex-ante efficient, i.e. it satisfies Conditions (9) and (10) (Lemma 1(i)), and ex-ante sustainable, i.e. it fulfills Condition (6) (Definition 2). With (A.33),(A.34) these conditions are

$$\begin{aligned} F\left(k, 1 - r\right) &= F(1, r) - k & \text{(A.35)} \\ F_R(1, r) F_K(k, 1 - r) &= F_R(k, 1 - r), & \text{(A.36)} \\ F(k, 1 - r) &= \overline{C}\ . & \text{(A.37)} \end{aligned}$$

There exists a unique value of $\overline{C} \in [0, \overline{C}^{EA,max}]$ so that this system can be solved for $(\tilde{r}, \tilde{k})$; with this value of $\overline{C}$, (A.35)–(A.37) reduce to Conditions (14),(15) and $(\tilde{r}, \tilde{k})$ is uniquely determined. To see this, note that $\overline{C}$ determines (r, k). Think of C_1 as a function of C_2 (defined by A.33, A.34, A.36, A.37 through variation of $\overline{C}$, where $\overline{C} = C_2$ as shown in Appendix A.1(i)) and consider first the minimal and maximal achievable consumption levels, indicated by $C_t^{EA,min}$ and $C_t^{EA,max}$, respectively. Setting $\overline{C} = 0$ implies $k^{EA,min} = 0$ and $r^{EA,min} = 1$, which yields $C_1^{EA,max} = F(1,1) > 0$ and $C_2^{EA,min} = F(0,0) = 0$. Setting $\overline{C} = \overline{C}^{EA,max}$ implies $k^{EA,max} = F(1, r^{EA,max})$. Inserting $k^{EA,max}$ into (A.36) uniquely yields $r^{EA,max}$, so that $C_1^{EA,min} = F(1, r^{EA,max}) - k^{EA,max} = 0$ and $C_2^{EA,max} = F(k^{EA,max}, 1 - r^{EA,max}) > 0$. By (A.37) we know that $dk/d\overline{C} = 1/F_K > 0$ and $dr/d\overline{C} = -1/F_R < 0$. As $F(\cdot, \cdot)$ is concave (by assumption) and C_1 is decreased linearly by increasing k as in (1), increasing k and reducing r by the ex-ante efficient mix (A.36) via increasing $\overline{C}$ from 0 to $\overline{C}^{EA,max}$ decreases C_1 monotonically from $C_1^{EA,max}$ to 0. As all functions involved are continuous $C_1(\overline{C})$ is continuous. Increasing $\overline{C}$ simultaneously increases C_2 continuously and monotonically from 0 to $C_2^{EA,max}$. Thus, by the intermediate value theorem and monotonicity, there exists a unique value of $\overline{C}$ so that the corresponding $(\tilde{r}, \tilde{k})$ yields $C_1 = C_2$, i.e. it fulfills (A.35).

A.4 Proof of Proposition 2

At time $t = 1b$, resource use $\tilde{r} = \tilde{R}_1$ is already irreversibly sunk in production and only the transfer of capital k can be adjusted. Note, that this analysis holds equally for utility as consumption is the only good in the model.

(i) As noted in Appendix A.3 $(\tilde{r}, \tilde{k})$ meets Condition (9) so $F_R(1, \tilde{r}) F_K(\tilde{k}, 1 - \tilde{r}) = F_R(\tilde{r}, 1 - \tilde{r})$. Thus, $(\tilde{r}, \tilde{k})$ cannot meet Condition (11) for ex-post efficiency as $F_R(1, \tilde{r}) F_K(\tilde{k}, 1 - \tilde{r}) - D'(\tilde{r}) F(\tilde{k}, 1 - \tilde{r})/(1 - D(\tilde{r})) < F_R(\tilde{k}, 1 - \tilde{r})$. Furthermore, $(\tilde{r}, \tilde{k})$ does not meet Condition (10) for ex-ante efficiency due to the damages $D(R_1)$. Thus, no adaptation of the ex-ante efficient and ex-ante sustainable policy $(\tilde{r}, \tilde{k})$ yields an allocation that is ex-ante and ex-post inefficient and ex-post unsustainable as

$$F(1,\tilde{r}) - \tilde{k} > (1 - D(\tilde{r}))F(\tilde{k}, 1 - \tilde{r}) \tag{A.38}$$

Increasing k to some $k^b > \tilde{k}$ to move towards a sustainable allocation does not allow to meet Condition (11). This is due to positive decreasing marginal utility in both inputs, which leads to: $F_R(1,\tilde{r})F_K(k^b, 1-\tilde{r}) - D'(\tilde{r})F(k^b, 1-\tilde{r})/(1-D(\tilde{r})) < F_R(1,\tilde{r})F_K(\tilde{k}, 1-\tilde{r}) - D'(\tilde{r})F(\tilde{k}, 1-\tilde{r})/(1-D(\tilde{r})) < F_R(\tilde{k}, 1-\tilde{r}) < F_R(k^b, 1-\tilde{r})$. Therefore, there exists no policy that is ex-post efficient and ex-post sustainable.

(ii) At time $t = 1b$ the social planner needs to resort to a higher capital transfer $k > \tilde{k}$ in her policy $(\tilde{r}, k)$ to achieve sustainability due to the damages $D(\tilde{r})$. As shown in (A.38) the second generation's consumption level is, $\hat{C}_2 = (1-D(\tilde{r}))F(k, 1-\tilde{r})$. With this condition for the behavior of the second generation and the irreversible production decision $\tilde{Y}_1 = F(1,\tilde{r})$ the effect of the level of capital transfer on utility can be derived

$$\text{for generation 1:} \quad \hat{C}_1 = F(1,\tilde{r}) - k \ , \tag{A.39}$$

$$\text{for generation 2:} \quad \hat{C}_2 = (1 - D(\tilde{r}))F(k, 1-\tilde{r}) \ . \tag{A.40}$$

The social planner sets k so that X is ex-post sustainable, i.e. it fulfills Condition (6) (Definition 2) and is on the RF-utility frontier, i.e. it fulfills Conditions (A.39) and (A.40). As the production function is concave an increase of k monotonically decreases the generation 1's consumption/utility level (A.39) while monotonically increasing the generation 2's consumption/utility level (A.40). Think of $C_1 - C_2$ as a function of k and consider minimal and maximal achievable consumption levels $C_t^{RF,min}, C_t^{RF,max}$ along the RF-utility frontier in (A.39) and (A.40) with irreversible resource inputs $\tilde{r} = \tilde{R}_1$ and $1 - \tilde{r} = \tilde{R}_2$. Setting $k = 0$ leads to $C_1^{RF,max} = F(1,\tilde{r}) > 0$ and $C_2^{RF,min} = (1-D(\tilde{r}))F(0, 1-\tilde{r}) = 0$. Setting $k = k^{RF,max}$ leads to, $C_1^{RF,min} = F(1,\tilde{r}) - k^{RF,max} = 0$ and $C_2^{RF,max} = (1-D(\tilde{r}))F(k^{RF,max}, 1-\tilde{r}) > 0$. As $F(K_t, R_t)$ is concave and C_1 is reduced linearly by increasing k, increasing k decreases $C_1 - C_2$ monotonically from $C_1^{RF,max} - C_2^{RF,min} > 0$ to $C_1^{RF,min} - C_2^{RF,max} < 0$. As all functions involved are continuous $C_1 - C_2$ is continuous. Thus, by the intermediate value theorem and monotonicity there exists a unique policy mix $(\hat{r}, \hat{k})$ with $\hat{r} = \tilde{r}$ that yields the allocation $\left(\hat{K}_1, \hat{R}_1, \hat{Y}_1, \hat{C}_1, \hat{K}_2, \hat{R}_2, \hat{Y}_2, \hat{C}_2\right)$ that is both on the RF-utility frontier and ex-post sustainable, i.e. if fulfills $C_1 = C_2$. Therefore, there exists a $\hat{k}$ for which $\hat{C_1} = F(1, \tilde{r}) - \hat{k} = (1 - D(\tilde{r}))F(\hat{k}, 1 - \tilde{r}) = \hat{C_2}$. As shown in Appendix A.4(i) for $\hat{k} > \tilde{k}$, Condition (11) for ex-post efficiency is not met.

(iii) As shown in Appendix A.4(i) $(\tilde{r}, \tilde{k})$ does not meet Condition (11) and $\tilde{R_1} = \tilde{r}$ is irreversible. At time $t = 1b$ k can still be adapted in the range $k \in [0, k^{EP,max} = F(1, \tilde{r})]$.

The social planner sets k so that X is ex-post efficient, i.e. it fulfills Conditions (11) and (12) (Lemma 1(ii)) and is on the RF-utility frontier, i.e. it fulfills (A.39) and (A.40). This leads to the following system:

$$F_R(1, \tilde{r})F_K(k, 1 - \tilde{r}) - D'(\tilde{r})F(k, 1 - \tilde{r})/(1 - D(\tilde{r})) = F_R(k, 1 - \tilde{r}), \tag{A.41}$$

$$C_2 = (1 - D(\tilde{r}))F\,(k, 1 - \tilde{r}) = \overline{C}\ . \tag{A.42}$$

There exists a unique value of $\overline{C} \in [0, \overline{C}^{EP,max}]$ so that this system can be solved for (r^*, k^*). To see this, note that $\overline{C}$ determines k and therefore C_2 in (A.40) and C_1 in (A.39). For ex-post efficiency consider the effect of minimal and maximal capital transfers on (A.41). From (A.41) define a function $\phi(k) = F_R(1, \tilde{r})F_K(k, 1 - \tilde{r}) - D'(\tilde{r})F(k, 1 - \tilde{r})/(1 - D(\tilde{r})) - F_R(k, 1 - \tilde{r})$. Setting $\overline{C} = 0$ implies $k^{EP,min} = 0$. For $\phi(0)$ this yields $\phi(0) = F_R(1, \tilde{r})F_K(0, 1 - \tilde{r}) - D'(\tilde{r})F(0, 1 - \tilde{r})/(1 - D(\tilde{r})) - F_R(0, 1 - \tilde{r}) > 0$. Setting $\overline{C} = \tilde{C}$ implies $\tilde{k}$ from Appendix A.3. As shown in Appendix A.4(i) this yields: $\phi(\tilde{k}) = F_R(1, \tilde{r})F_K(\tilde{k}, 1 - \tilde{r}) - D'(\tilde{r})F(\tilde{k}, 1 - \tilde{r})/(1 - D(\tilde{r})) - F_R(\tilde{k}, 1 - \tilde{r}) < 0$.

As a decreasing k monotonically increases $F_K(k, 1 - \tilde{r})$, monotonically decreases $F(k, \tilde{r})$ and monotonically decreases $F_R(k, 1 - \tilde{r})$, $\phi(k)$ is monotonically decreasing in k. Varying k from 0 to $\tilde{k}$ changes $\phi(k)$ from $\phi(0) > 0$ to $\phi(\tilde{k}) < 0$. As all functions involved are continuous $\phi(k)$ is continuous. Thus, by the intermediate value theorem and monotonicity, there exists a unique value of k and a corresponding value of $\overline{C}$ so that $\phi(k^*) = 0$ and (r^*, k^*) is ex-post efficient, i.e. it fulfills (A.41). Therefore, there exists a unique policy (r^*, k^*) with $r^* = \tilde{r}$ and $k^* < \tilde{k}$ that yields an allocation $X^* = (K_1^*, R_1^*, Y_1^*, C_1^*, K_2^*, R_2^*, Y_2^*, C_2^*)$ that is ex-post efficient.

A.5 Proof of Proposition 3

When the social planner sets (r, k) at time $t = 1a$ under awareness of the damages $D(R_1)$ the generations' consumption levels at time $t = 1b$ are:

$$C_1 = F(1, r) - k \ , \tag{A.43}$$

$$C_2 = (1 - D(r))F(k, 1 - r) \ . \tag{A.44}$$

Note, that this analysis holds equally for utility as consumption is the only good in the model. For an ex-post efficient allocation Conditions (11) and (12) must hold (Lemma 1(ii)), and Condition (6) for sustainability (Definition 2). With (A.43) and (A.44) these conditions are:

$$(1 - D(r))F(k, 1 - r) = F(1, r) - k \tag{A.45}$$

$$F_R(1, r)F_K(k, 1 - r) - D'(r)F(k, 1 - r)/(1 - D(r)) = F_R(k, 1 - r), \tag{A.46}$$

$$(1 - D(r))F(k, 1 - r) = \overline{C} \ . \tag{A.47}$$

There exists a unique value of $\overline{C} \in [0, \overline{C}^{Bliss,max}]$ so that this system can be solved for (r, k); with this value of $\overline{C}$ (r, k) is uniquely determined. To see this, note that $\overline{C}$ determines (r, k). Think of C_1 as a function of C_2 (defined by A.43, A.44, A.46, A.47 through variation of $\overline{C}$, where $\overline{C} = C_2$ as shown in Appendix A.1(i)) and consider first the minimal and maximal achievable consumption levels, indicated by $C_t^{Bliss,min}$ and $C_t^{Bliss,max}$, respectively. Setting $\overline{C} = 0$ implies $k^{Bliss,min} = 0$ and $r^{Bliss,min} = 1$, which yields $C_1^{Bliss,max} = F(1, r^{Bliss,min}) > 0$ and $C_2^{Bliss,min} = (1 - D(r^{Bliss,min}))F(0, 1 - r^{Bliss,min}) = 0$. Setting $\overline{C} = \overline{C}^{Bliss,max}$ implies $k^{Bliss,max} = F(1, r^{Bliss,max})$. Inserting $k^{Bliss,max}$ into Equation (A.46) uniquely yields $r^{Bliss,max}$, so that $C_1^{Bliss,min} = F(1, r^{Bliss,max}) - k^{Bliss,max} = 0$ and $C_2^{Bliss,max} = (1 - D(r^{Bliss,max}))F(k^{Bliss,max}, 1 - r^{Bliss,max}) > 0$. By (A.47) we know that $dk/d\overline{C} = 1/(1 - D(r))F_K > 0$ and $dr/d\overline{C} = -1/((1 - D(r))F_R - D'F(k, 1-r)) < 0$. As F is concave and C_1 is decreased linearly by increasing k as in (1), increasing k and reducing r by the ex-post efficient mix (A.46) via increasing $\overline{C}$ from 0 to $\overline{C}^{Bliss,max}$ decreases C_1 monotonically from $C_1^{Bliss,max}$ to 0. As all functions involved are continuous $C_1(\overline{C})$ is continuous. Increasing $\overline{C}$ simultaneously increases C_2 continuously and monotonically from 0 to $C_2^{Bliss,max}$. Thus, by the intermediate value theorem and monotonicity, there exists a unique value of $\overline{C}$ so that the corresponding (r^{Bliss}, k^{Bliss}) and allocation $X^{Bliss} = \left(K_1^{Bliss}, R_1^{Bliss}, Y_1^{Bliss}, C_1^{Bliss}, K_2^{Bliss}, R_2^{Bliss}, Y_2^{Bliss}, C_2^{Bliss},\right)$ ensure $C_1 = C_2$, i.e. fulfill (A.45).

This allocation is reduced feasible, that is feasible under irreversibility and unawareness as: at $t = 1a$ the social planner can choose any $r \in [0, 1]$. From the existence proof its clear that $r^{Bliss} \in [0, 1]$. So, in $t = 1a$ the social planner chooses r^{Bliss} and some matching k. At $t = 1b$ the damage becomes apparent and r^{Bliss} is fixed. Still, k can be adjusted $k \in [0, F(1, r^{Bliss})]$ generating a RF-utility frontier as in the existence proof. Therefore, the policy (r^{Bliss}, k^{Bliss}) which yields an ex-post efficient and ex-post sustainable allocation is reduced feasible, that is feasible under unawareness and irreversibility.

Paper 4:
Merit good arguments in sustainability discussions: Challenges and opportunities

Merit good arguments in sustainability discussions: Challenges and opportunities

Nikolai Hoberg[*]

Department of Sustainability Science and Department of Economics,
Leuphana University of Lüneburg, Germany

Abstract:
Merit goods are often invoked in situations where conclusions from welfare economic analysis seem ethically unappealing or incomplete in cases such as education, arts and health care. In this paper, I clarify how merit good arguments deviate from individual preferences by following Besley's model (1988). Building on Goodin (1989), I relate merit good arguments and the justification for deviations from individual preferences to two conceptions of well-being: an informed preference satisfaction and a perfectionist conception. Building on this framework, I outline several challenges and opportunities that result from applying merit good arguments in intergenerational contexts.

JEL-Classification: D63, H23, Q56

Keywords: merit goods, sustainability, well-being

[*] Corresponding author: Sustainability Economics Group, Leuphana University of Lüneburg, P.O. Box 2440, D-21314 Lüneburg, Germany, phone: +49.4131.677-2715, fax: +49.4131.677-1381, email: hoberg@uni.leuphana.de.

4.1 Introduction

Many governments in the world subsidize social housing, primary and secondary education, and healthcare. In this context, merit good arguments are often invoked when conclusions from welfare economic analyses seem not to justify these at first glance ethically appealing measures. Unfortunately, in these examples merit good arguments for government intervention are difficult to disentangle from other arguments for government intervention such as the ones based on externalities or distributive concerns over equitable allocations. The concept of merit goods was introduced in two contributions by Richard Musgrave (1957, 1959). In his article for the New Palgrave Dictionary of Economics, he describes merit goods, somewhat vaguely, as a situation "where evaluation of a good (its merit or demerit) derives not simply from the norm of consumer sovereignty but involves an alternative norm" (Musgrave 1987: 579). Thus, merit good arguments posit preferences that deviate from individual preferences due to some 'alternative norm' for some specific good. That is, merit goods arguments justify government intervention on education, arts and healthcare in an allocation that is Pareto-efficient and equitable, i.e. without distributive concerns, through a deviation from individual preferences (Andel 1984).

With this deviation from individual preferences the concept of merit goods poses fundamental questions regarding the ethical foundation of welfare economics. In particular, merit good arguments lead to two questions: The first question concerns the conception of well-being in merit goods, i.e. why do merit goods make people better-off beyond their contribution to individual utility? The second question concerns the justification of paternalist policies, i.e. why are government policies that are not based on individual preferences justified?

While these questions have been extensively debated in the public finance literature, they are also important for discussions of sustainability. The first question is relevant for sustainability criteria that compare future and present well-being. For example, one common criterion states that the future should not be worse-off than the present, i.e. that utility should be non-decreasing over time (Pezzey 1997). Then the question becomes whether and how merit good arguments and their conceptions of well-being allow such comparisons of well-being between future and present generations. The second question is also interesting to sustainability as it involves the justification of government intervention based on normative objectives that are not derived from individual preferences. That is, insights on the justification of paternalist policies from the merit good literature illustrate more general arguments on the pursuit of normative objectives in economic analysis.

In this article, I (i) clarify the definition of merit goods by considering the model for merit good arguments in Besley (1988), (ii) I build on the work of Goodin (1989) and discuss two conceptions of well-being that can serve as an underpinning for merit good arguments, and (iii) discuss what challenges and opportunities merit good arguments raise vis-a-vis the discussion of sustainability problems.

The paper is structured as follows: Section 4.2 discusses the history of merit goods, provides a definition of the term and introduces a simple economic framework to give a more analytical structure to the argument. Section 4.3 examines conceptions of well-being behind merit good arguments and associated justifications for paternalist policies. Section 4.4 discusses the challenges and opportunities of merit goods in an intergenerational framework and sustainability discussions. Section 4.5 concludes the discussion.

4.2 History and definition of merit goods

Richard Musgrave introduced the concept of merit goods in two early contributions in 1957 and 1959. Despite the long history and broad use of the concept since then, confusion around the definition of the concept remains. In his article for the New Palgrave Dictionary of Economics, he concludes that "[i]n all, it seems difficult to assign a unique meaning to the term" (Musgrave 1987: 581). Andel (1984) traces the development and changes of the definition of merit goods in Musgrave's works. He criticizes two aspects in Musgrave's use of the concept. First, the given examples involve arguments for government intervention that are independent from merit goods: externalities, which justify Pigouvian taxes, and redistribution, which justifies redistributive taxation.

Therefore, the examples do not serve particularly well to explicate the specific argument for government action behind merit goods. Second, he criticizes that Musgrave's use of the concept changes between a normative theory, that people should consume higher amounts of merit goods, to a positive one, that explains why governments subsidize specific merit goods, e.g. due to paternalist altruism that targets specific goods (Andel 1984: 637).

Andel concludes that due to the confusion around the concept, merit goods are mostly used for the case of a social planner who increases the consumption of some good due to a variety of arguments (Andel 1984: 648). This leads us to define merit goods as:

Definition 1 (Merit good)
A *merit good* is a good for which a social planner prescribes preferences different from individual preferences.

This definition distinguishes merit goods on the one hand from externalities and public goods and on the other from distributive concerns. Further, it says that a social planner can disregard individual preferences in case of a merit good.

Despite this vague definition, merit good arguments have made their way into some parts of microeconomic theory such as optimal taxation (e.g. Besley 1988, Schroyen 2005, Blomquist and Micheletto 2006, O'Donoghue and Rabin 2006, Schroyen 2010) and the discussion on paternalism in behavioral economics (e.g. Schnellenbach 2012). Also, in his discussion of the normative foundation of economic analysis, Partha Dasgupta refers to merit goods as an established concept for situations where policies are justified independently from individual preferences (Dasgupta 2005). In this way, merit goods are invoked in many applied areas: in subsidies for organic farming (Mann 2003), water supply and sanitation (Schwartz and Schouten 2007), the role of social impact bonds in health care (Fitzgerald 2013), subsidies on art and culture (Soh 2011), housing subsidies (ter Rele and van Steen 2003) or sin-taxes on alcohol or sugar to fight unhealthy lifestyles (e.g. O'Donoghue and Rabin 2006, Schroyen 2010).

In order to give more structure to Definition 1, I will use the model for merit goods in Besley (1988) for clarification. For this, consider an economy with two goods and one individual i where x_i is the amount of a normal good and y_i the amount of a merit good consumed by individual i. Assume a *utility function* of individual i, based on its actual preferences, with positive decreasing marginal utility in both goods:

$$U_i(x_i, y_i) \tag{1}$$

In welfare economics, this utility function usually is taken to reflect individual choice as well as an index for the evaluation of individual well-being (Hausman and McPherson 1993: 680). This means that according to U_i every individual choice makes the individual better-off.

With merit goods a social planner respects the individuals choices for good x_i, but prescribes different preferences for the merit good y_i. In Besley's (1988) model, this difference leads to, what I call, the *merit utility function* of individual i, based on merit preferences from a social planner for good y_i.

$$V_i(x_i, y_i) := U_i(x_i, \mu(y_i)) \tag{2}$$

Also, the merit good function $\mu(y_i)$ is assumed to be strictly concave so that positive decreasing marginal utility in both goods is maintained. Here, V_i serves as an index for the evaluation of the well-being of individual i from the perspective of the social planner. It does not reflect the actual choices made by individual i. That is, it can be that the individual prefers an allocation, i.e. $U_i(x_i^a, y_i^a) < U_i(x_i^b, y_i^b)$, yet that this choice actually decreases merit utility, i.e. $V_i(x_i^a, y_i^a) > V_i(x_i^b, y_i^b)$. Besley further specifies a merit utility function with $\mu(y_i) = \theta y_i$ which leads to $V_i(x_i, y_i) := U_i(x_i, \theta y_i)$. This scaling model allows to easily distinguish merit goods with $\theta > 1$ from demerit goods with $\theta < 1$.[1]

As the social planner's evaluation deviates from individual choice, that is, V_i is not the same function as U_i, this might lead the social planner to correct peoples choices

[1] This specific scaling approach to model merit goods has been criticized in Schroyen (2005) for leading to counter-intuitive cases where under optimal taxation a merit good should be taxed and the normal good subsidized if demand for merit goods is inelastic. Yet, for the purpose of conceptual clarification Besley's model is fine.

with regard to merit goods. For this she could use different instruments such as taxes and subsidies on merit goods. Besley derives a condition for optimal merit good taxes $t = (1 - \theta)/\theta$ (Besley 1988: 376, Equation 3.8). It leads to the obvious result that (de-)merit goods $\theta > 1$ ($\theta < 1$) should be subsidized (taxed).[2,3]

In sum, this model allows to disentangle externalities, distribution and merit goods. Merit good arguments concern the form of merit utility and $\mu(y_i)$, i.e. the deviation from individual preferences for some good. For example, this could justify to 'merit taxes' on cigarettes to improve peoples health. Externalities concern external effects that originate, for example, from the consumption of good y_i. For example, this could justify Pigouvian taxes in order to internalize external damages from smoking to others peoples health. Similarly, y_i could be a public good in this model (as discussed in Andel 1984). Redistribution in this model concerns inequality of different levels of merit utility V_i. For example, there could be transfers between individuals with high or low levels of merit utility. Thus, this model clarifies the specific economic theory behind merit good arguments.

With the background of this analytical framework, I now discuss some questions on the ethical justification of taxes and subsidies on merit goods in the next section.

4.3 Conceptions of well-being and paternalist policies

As merit good arguments deviate from the exclusive reliance on individual preferences, they pose many philosophical questions regarding moral reasoning in economic analysis and beyond. One common criticism is that merit goods are an instance of paternalism (e.g. Schnellenbach 2012). Dworkin (2010: Sec. 1) defines paternalism as:

[2] Besley also considers the case where preferences are heterogenous between individuals, which requires individual weights θ_i and individual merit taxes.

[3] Regarding alternative instruments, Andel notes that Musgrave never considers the provision of information for merit goods as an instrument (Andel 1984).

Definition 2

A policy is *paternalist*, if it is justified on the grounds that the subjected person would be better-off due to this policy and would not consent to be treated this way.[4]

Following this definition paternalism raises two questions regarding merit goods which I will briefly discuss. First, what is the conception of well-being in merit goods, i.e. why do merit goods make people better-off beyond their contribution to individual utility? Second, why are government policies that are not based on individual preferences justified?

In a first step, consider two common conceptions of well-being regarding the utility function U_i (as in Hausman and McPherson 2006: 118). The first is a *hedonist conception* that says that if an individual chooses an allocation over another $U_i(x_1^a, y_1^a) < U_i(x_1^b, y_1^b)$, then this choice increases the happiness experienced by the individual, i.e. the latter allocation leads to a mental state with more happiness. A common criticism of this conception, according to Hausman and McPherson, is that it would allow to put people in an 'experience machine' where the mental state of happiness is created, but their desires are not actually satisfied. This experience machine could create "intense physical pleasures or experiences of climbing Everest" (Hausman and McPherson 2006: 122), without individuals being aware they are not actually fulfilling these preferences. This criticism leads to the second conception that Hausman and McPherson consider: the *preference satisfaction conception.* It says that if an individual chooses an allocation over another $U_i(x_1^a, y_1^a) < U_i(x_1^b, y_1^b)$, then this choice is good if the individuals preferences or desires are actually satisfied through this choice. This conception, therefore, rules out the experience machine which only creates the mental state of happiness rather than the actual satisfaction of preferences. Still, this conception leaves the determination of preferences entirely up to the individual, as does the hedonist conception. Thus, both conceptions do not require paternalist policies as individuals always make the right

[4] The literal definition of paternalist policies in Dworkin is as follows: "When they are justified solely on the grounds that the person affected would be better off, or would be less harmed, as a result of the rule, policy, etc., and the person in question would prefer not to be treated this way, we have an instance of paternalism"(Dworkin 2010: Sec. 1).

choice in either increasing their happiness or the satisfaction of their preferences.

If a social planner formulates a merit utility function V_i, then this rules out the hedonist and the actual preference satisfaction conception of well-being due to the prescription of preferences for the merit good. Goodin (1989) discusses merit good arguments in relation to an 'utilitarian' conception of well-being and a perfectionist conception of well-being. Both are discussed in a democratic setting where there is democratic accountability, some form of representative democracy, and sustained political support for public policies (Goodin 1989: 254). Here, I will follow his discussion with small variations and highlight connections to Musgrave's writings.

The first conception that I consider for the merit utility function takes preference satisfaction a step further by considering *'informed' or 'rational' preference satisfaction* (Hurka 2006: 363). This says that if an individual prefers an allocation over another, i.e. $U_i(x_1^a, y_1^a) < U_i(x_1^b, y_1^b)$, then this increases its well-being only if the underlying preference is well-informed and rational. This would allow the social planner to use function V_i not as a description of individual choice, but rather an external index of informed preference satisfaction. That is, if the individual were better informed it would choose differently between the two allocations, i.e. $V_i(x_1^a, y_1^a) > V_i(x_1^b, y_1^b)$.[5]

While Goodin generally speaks of a 'utilitarian' conception of well-being, his 'retrospective rationality' justification for paternalism fits well with the conception of informed preference satisfaction. The idea is that "your subsequent, settled preferences will ratify the present choices I make on your behalf but against your current will" (Goodin 1989: 257). That is, temporary paternalist policies are justified insofar as people can either subsequently agree to this policy or disagree and abolish it in democratic elections. This justification is very similar to what Musgrave mentions with regard to education as a merit good. Here, he says the benefits of "education are more evident to the informed than the uninformed, thus justifying the compulsion in the allocation of resources to education" (Musgrave 1959: 14). Further, he refers to the safeguard of democratic leadership: "While consumer sovereignty is the general rule, situations may arise, in the context of a democratic society, where an informed group is justified in imposing its

[5] This distinction also comes up in Harsanyi's (1982) distinction of 'manifest' and 'true preferences'.

decision on others" (Musgrave 1959: 14).

Two other justifications[6] that Goodin links to a 'utilitarian' conception of well-being are derived from implicit choices of individuals. The 'implicit consumer choice' justification says that paternalist policies are justified if merit goods are universal means that "can be tied to the natural implications of choices people actually have made" (Goodin 1989: 248). This means that people might not demand merit goods such as education directly, but that their later choices in life based on this good justify government provision of merit goods such as education. The 'implicit political choice' justification says that paternalist policies are justified if merit goods are "implicit within the actual political choices people make"(Goodin 1989: 249). For example, this means that if people participate in a democratic election, implicitly want everyone "to be fairly fully-informed about the issues they are voting on"(Goodin 1989: 249). Both of these justifications also allow the use of V_i as an implicitly justified external index for some goods.

The second conception is a *perfectionist conception* which "identifies states of affairs, activities, and/or relationships as good in themselves and not good in virtue of the fact that they are desired or enjoyed by human beings" (Wall 2012: Introduction). Goodin himself considers a central argument for merit goods as "manifestations [...] of some form of perfectionism" (Goodin 1989: 244). This says that if an individual consumes one bundle rather than another, i.e. $V_i(x_1^a, y_1^a) > V_i(x_1^b, y_1^b)$, then this increases the attainment of values such as knowledge or aesthetic pleasure irrespective of individual preferences. Thus, perfectionism allows the social planner to use V_i as an external index for the attainment of perfectionist objectives. Further, no amount of information or experience with the merit good will necessarily lead the individual in this conception to prefer it over other goods.

[6] Goodin considers two further justifications that relate to a utilitarian conception: one based on paternalist altruism, i.e. altruism that targets specific goods, that he calls 'externalities'; and a second based on minority moral judgements, e.g. moral judgements from minority religious groups, that he calls 'high moralism'. I do not discuss these here as they do not provide justifications pertinent to this discussion.

The justification Goodin discusses for paternalist policies based on perfectionism relies on the fact that peoples 'political preferences' can be different from their market preferences. He says that what people are "doing when engaging in such political action is pursuing some vision of the public good, as a matter of moral principle, rather than pursuing private self-interest of any sort" (Goodin 1989: 254). This is again close to Musgrave's argument on the safeguard of democratic leadership. Similarly, Musgrave refers to 'community values' where "[i]ndividuals, as members of the community, accept certain community values or preferences, even though their personal preferences might differ" (Musgrave 1987: 580). This justification says that people can vote for the provision of merit goods such as education based on their political preferences against their market preferences as in U_i. In this vein, Brennan and Lomasky show that merit goods can be provided publicly in case preferences revealed in political arena are different from those revealed in the market place (Brennan and Lomasky 1984).

As a side note, Goodin mentions the support for environmental policies that do not serve peoples self-interest as an example for the 'political preferences' justification. This justification is close to what other authors in the environmental field discuss under different conceptions of the human being such as homo politicus vs. homo oeconomicus (e.g. Faber, Petersen and Schiller 2002).

In sum, we saw there are different justifications for paternalist policies derived from a preference satisfaction conception of well-being ('retrospective rationality', 'implicit consumer choice', 'implicit political choice') and a perfectionist conception of well-being ('political preferences'). All of these justifications rely on some form of democratic accountability or directly observable acts. These justifications were laid out later, and significantly completed by other authors. The confusion surrounding merit goods might, therefore, also be due to the comparatively imprecise ethical justification of merit goods that Musgrave himself provided.

4.4 Challenges and opportunities for sustainability discussions

In Section 4.3, we saw that merit good arguments are connected to specific conceptions of well-being. In the following, I discuss what merit good arguments can bring to discussions of sustainability. I consider five points that either pose challenges or opportunities for the economic analysis of sustainability problems.

The first point is directly related to the role of conceptions of well-being in theories of justice. This is relevant if sustainability is defined in terms of intra- and intergenerational justice (e.g. Baumgärtner and Quaas 2010). As a theory of justice is a complex concept it is useful to consider three minimum 'bases of distinction' which theories of justice conceptualize differently (see Sen 2008). These bases are (i) a metric of individual advantage such as utility or capabilities, (ii) an aggregation rule over this metric such as summation in utilitarianism or maximin, and (iii) the priority of some aspect of individual advantage over others such as the strict priority of basic liberties over considerations for inequality in Rawls difference principle (Sen 2008: Sec. 7). Here, conceptions of well-being and, therefore, the debate on merit goods is relevant for (i) the metric of individual advantage in a theory of justice. They are not relevant for (ii) aggregation rules or (iii) the priority of some aspect of individual advantage. Thus, the definition of sustainability as inter- and intragenerational justice automatically leads to the question what metric of individual advantage and what conception of well-being should be used. This also involves the question of how this metric incorporates or deviates from individual preferences. That is, when sustainability is defined in terms of justices then this automatically either raises or avoids merit good arguments depending on the metric of individual advantage. Thus, if sustainability involves conceptions of well-being that go against individual preferences, then the analytical results on merit good arguments can provide an opportunity for discussions of sustainability problems.

The second point relates to how welfare economics incorporates questions of justice in social welfare functions. To relate this point to the model from above, consider two generations $i = 1, 2$ and a representative individual for each generation with utility functions U_1, U_2 and merit utility functions V_1, V_2. This allows the formulation of the following intergenerational *social welfare function*:

$$W(U_1, U_2) \tag{3}$$

This definition of a social welfare allows different aggregation rules based on the metric of individual utility – as under (ii) in the second basis of distinction above. For example, such social welfare functions are applied in sustainability discussions such as the calculation of the benefits and costs of climate change (Stern 2007). There exists a broad literature on how intergenerational justice can be incorporated in social welfare functions (e.g. Asheim 1988, Chichilnisky 1996, Alvarez-Cuadrado and Van Long 2009). One common criticism of such intergenerational social welfare functions is that they are based on the assumption that future preferences, here U_2, are known and comparable which has lead some to consider opportunity based measures of intergenerational justice (e.g. Howarth 1997). Still, for conceptual discussions on intergenerational justice this is usually taken as granted.

As discussed in Section 4.2 merit good arguments lead to merit utility as the metric of individual advantage. If this metric is assumed for both generations, this leads to the formulation of a *merit social welfare function*:

$$W(V_1, V_2) \tag{4}$$

The definition of social welfare with the metric of merit utility also allows different aggregation rules. This, of course, follows from the difference between the concept of merit goods and distributive concerns as discussed in Section 4.2. In this sense, merit good arguments are clearly distinct from discussions on intergenerational distribution. That is, merit good arguments still leave the challenge what social welfare function best captures notions of intergenerational justice.

The third point relates to the informational requirements of different conceptions of well-being. The usual formulations of intergenerational social welfare functions take the form as in Equation (3). One problem with the formulation as Equation (4) is the deviation from individual preferences requires further information to be justified. That is, all of the justifications for the deviation from individual preferences in Section 4.3 rely on some form of democratic accountability and information on outcomes of political processes (as in Goodin 1989: 251). For example, the deviation from individual preferences in the perfectionist conception of well-being relies on the justification via 'political preferences'. Thus, using V_2 for the second generation presupposes knowledge, not only of future preferences, but also on the outcome of democratic processes. In this regard, merit good arguments pose a particular challenge in sustainability discussions due to their higher informational requirements for the justification of deviations from individual preferences. An associated challenge lies in the question how merit utility is comparable between generations. For example, one could argue that the utility function of a representative individual is the same for both generations, i.e. $U_i(x, y)$ for $i = 1, 2$, but that different weights θ_i on a merit good or different merit goods are determined, for example, due to differing 'political preferences'. This then leads to the challenge of how well-being can be compared between generations when there are different merit utility functions, i.e. $V_1(x, y)$ is not the same function as $V_2(x, y)$.

The fourth point pertains to the case of externalities and paternalist policies associated with merit good arguments. It could be interesting to use a social welfare function as in Equation (4) in a case where there is an intergenerational externality originating from a merit good. Here, the implementation of paternalist policies in the first generation such as merit taxes could have a positive effect on well-being of the second generation due to the reduction of negative intergenerational externalities. For example, if one wants to reduce obesity rates and implements merit taxes on sugar, then this could also reduce potential externalities from sugar plantations and agricultural production such as biodiversity loss, land degradation and the like. In this sense, merit good arguments provide the opportunity to study the interaction between different arguments for government intervention on one particular good, namely Pigouvian taxes due to the objective of Pareto-efficiency and merit taxes due to the objective of merit utility.

The fifth and final point concerns uncertainty regarding the determination of future merit goods. The question is how to incorporate future merit utility in a merit social welfare function such as Equation (4) when there are several potential goods the future generation might determine as merit goods. In this case there is not a predetermined function V_2 that can be used in Equation (4), but rather the need to provide the opportunity set of the future generation to determine their merit goods by themselves. This leads to the opportunity that merit good arguments and associated conceptions of well-being require the formulation of novel sustainability constraints in intergenerational models.

4.5 Conclusion

In this article, I discussed the origin of the concept of merit goods and its current application in economic analysis. While merit good arguments are invoked in many contexts, we saw that their definition and ethical justification remains quite vague. For this, I discussed different conceptions of well-being that can serve as an ethical underpinning for merit good arguments that lead to deviations from individual preferences for some goods. We saw that an informed preference satisfaction and a perfectionist conception of well-being came close to what Musgrave mentioned in some of his writings. Following Goodin (1989), we saw that both conceptions justify policies that deviate from individual preferences through some variant of democratic accountability – which is similar to what Musgrave argued initially.

As to the question of what merit good arguments bring to sustainability discussions, I outlined several challenges and opportunities that the justifications for the deviation from individual preferences raise in an intergenerational context. For example, we saw that these justifications require additional information, for example, on ‘political preferences’ to justify the deviation from individual preferences. Further, we also saw that merit good arguments do not answer the challenge which social welfare function best captures intergenerational justice and that merit good arguments are clearly distinct from questions of intergenerational distribution.

Merit good arguments are currently not as prominently discussed as other approaches in economics that have deviated from individual preferences. This might be due to the vague definition of merit goods and that these other approaches provide more specific ethical justifications. For example, Amartya Sen's capability approach details why individual preferences can be a problematic guide in some contexts (Sen 1980) and how the capability approach can make a difference in development policy (Sen 1999). Similarly, the discussion on distributive justice shows how well-justified non-welfarist philosophical concepts such as equality in resources or primary goods can be brought into economic theory (e.g. Roemer 1996). The opportunity associated with merit good arguments in this context is that they provide analytical results for the case when there is a difference between what an individual does and what is considered good for its well-being – in contrast to the common assumption that any free individual action increases individual well-being. This connects to the ongoing debate on the normative foundation of economic analysis (e.g. Sen 1977, Hausman and McPherson 2006) and the potential gains in a systematic discussion on conceptions of well-being between ethics and economics. It has hopefully become clear in this paper that the debate on conceptions of well-being and merit good arguments can be productive for sustainability problems, for example, if it concerns the role of environmental goods in well-being and eases or hinders the formulation of theories of intergenerational justice in economic theory.

Acknowledgements

I am grateful to Dave Abson, Stefan Baumgärtner, John-Oliver Engler, Joachim Fünfgelt and Sergio Mukherjee for critical and constructive discussion and the German Federal Ministry of Education and Research for financial support under grant 01UN1011A.

Bibliography

Alvarez-Cuadrado, F. and Van Long, N. (2009). A mixed Bentham–Rawls criterion for intergenerational equity: Theory and implications. *Journal of Environmental Economics and Management*, 58(2):154–168.

Andel, N. (1984). Zum Konzept der meritorischen Güter. *FinanzArchiv*, New Series, 42(3):630–648.

Asheim, G. (1988). Rawlsian intergenerational justice as a Markov-perfect equilibrium in a resource technology. *Review of Economic Studies*, 55(3):469–483.

Baumgärter, S. and Quaas, M. (2010). What is sustainability economics? *Ecological Economics*, 69:445–450.

Besley, T. (1988). A simple model for merit good arguments. *Journal of Public Economics*, 35(3):371–383.

Blomquist, S. and Micheletto, L. (2006). Optimal redistributive taxation when government's and agents' preferences differ. *Journal of Public Economics*, 90(6-7):371–383.

Brennan, G. and Lomasky, L. (1984). Institutional aspects of merit good analysis. *FinanzArchiv*, New Series, 41(2):183–206.

Chichilnisky, G. (1996). An axiomatic approach to sustainable development. *Social Choice and Welfare*, 13(2):231–257.

Dasgupta, P. (2005). What do economists analyze and why: values or facts? *Economics and Philosophy*, 21(2):221–278.

Dworkin, G. (2010). Paternalism. In: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2010 Edition, [Online available at: `http://plato.stanford.edu/archives/sum2010/entries/paternalism/`]

Faber, M., Petersen, T., and Schiller, J. (2002). Homo oeconomicus and homo politicus in Ecological Economics. *Ecological Economics*, 40(3):323–333.

Fitzgerald, J. (2013). Social impact bonds and their application to preventive health? *Australian Health Review*, 37(2):199–204.

Goodin, R. (1989). Stars to steer by: the political impact of moral values. *Journal of Public Policy*, 9(3):241–259.

Harsanyi, J. (1982). Morality and the theory of rational behaviour. In: Sen, A. and Williams, B. (eds.), *Utilitarianism and beyond*, Cambridge University Press, Cambridge. 39–62.

Hausman, D. and McPherson, M. (1993). Taking ethics seriously: Economics and contemporary moral philosophy. *Journal of Economic Literature*, 31(2):671–731.

Hausman, D. and McPherson, M. (2006). *Economic Analysis, Moral Philosophy, and Public Policy*, 2nd edition. Cambridge University Press, Cambridge.

Howarth, R. (1997). Sustainability as opportunity. *Land Economics*, 73(4):569–579.

Hurka, T. (2006). Value theory. In: Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, Oxford. 357–359.

Mann, S. (2003). Why organic food in Germany is a merit good. *Food Policy*, 28(5-6):459–469.

Musgrave, R. (1957). A multiple theory of budget determination. *FinanzArchiv*, New Series, 17(3):333–343.

Musgrave, R. (1959). *Theory of Public Finance.* McGraw Hill, New York.

Musgrave, R. (1987). Merit goods. In: Eatwell, J., Milgate, M. and Newman, P. (eds), *The New Palgrave: A Dictionary of Economics*, Vol. 4, Macmillan, London. 792–793.

O'Donoghue, T. and Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, 90(10-11):1825–1849.

Pezzey, J. (1997). Sustainability constraints versus "optimality" versus intertemporal concern, and axioms versus data. *Land Economics*, 73(4):448–466.

Rawls, J. (1971). *A Theory of Justice.* Harvard University Press, Cambridge, Mass.

Roemer, J. (1996). *Theories of Distributive Justice.* Harvard University Press, Cambridge, Mass.

Schroyen, F. (2005). An alternative way to model merit good arguments. *Journal of Public Economics*, 89 (5-6):957–966.

Schroyen, F. (2010). Operational expressions for the marginal cost of indirect taxation when merit arguments matter. *International Tax and Public Finance*, 17(1):43–51.

Schnellenbach, J. (2005). Nudges and norms: on the political economy of soft paternalism. *European Journal of Political Economy*, 28(2):266–277.

Schwartz, K. and Schouten, M. (2007). Water as a political good: revisiting the relationship between politics and service provision. *Water Policy*, 9(2):119–129.

Sen, A. (1977). Rational fools: a critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, 6(4):317–344.

Sen, A. (1980). Equality of what? In: McMurrin, S. (ed), *The Tanner Lecture on Human Values*, Vol. 1, Cambridge University Press, Cambridge. 197-220.

Sen, A. (1999). *Development as Freedom.* Oxford University Press, Oxford.

Sen, A. (2008). Justice. In: Durlauf, S., and Blume, L. (eds), *The New Palgrave: A Dictionary of Economics*, 2nd edition, Macmillan, London. 792–793.

Soh, B. (2011). Lambert Zuidervaart: Art in public: politics, economics, and a democratic culture. *Journal of Cultural Economics*, 36:87–90.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review.* Cambridge University Press, Cambridge.

ter Rele, H. and van Steen, G. (2003). Measuring housing subsidies: Distortionary and distributional effects in the Netherlands. *Fiscal Studies*, 24(3):317–339.

Wall, S. (2012). Perfectionism in moral and political philosophy. In: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2012 Edition, [Online available at: `http://plato.stanford.edu/archives/win2012/entries/perfectionism-moral/`]