Hoberg, N., **Strunz, S.** (2018):
When individual preferences defy sustainability - Can merit good arguments close the gap?
Ecol. Econ. **143** , 286 – 293

http://dx.doi.org/10.1016/j.ecolecon.2017.07.004

When individual preferences defy sustainability - can merit good arguments close the gap?

Abstract:

In this paper, we discuss how merit good arguments may contribute to discussions about sustainability. To this end, we clarify how merit good arguments deviate from individual preferences and relate 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, we analyze how merit good arguments can be helpful for discussing sustainability as justice, what challenges merit good arguments pose to future generations, and whether they can serve as a normative justification for green nudges. The analysis yields two main insights. First, a reflection on the concept of merit goods is helpful in sorting out the different justifications that sustainability interventions may rely on. In particular, it allows separating the challenges of redistribution, internalization of externalities and increasing individual consumption of particular (merit) goods such as health care or education more clearly. Second, the precise notion of merit goods by itself, however, only offers a limited contribution and does not represent a blank check to justify deviations from individual preferences.

JEL-Classification: D63, H23, Q56

Keywords: merit goods, sustainability, well-being

1 Introduction

Sustainability, defined in terms of intra- and intergenerational justice (e.g. in Baumgärtner and Quaas 2010) goes beyond the internalization of externalities. Moreover, individual preferences need not necessarily facilitate sustainability, which leads to the question of how this gap can be bridged (e.g. Pezzey 2004, Bullock and Collier 2011). Thus, one crucial issue for sustainability economics consists in reflecting on the appropriate scope of and justification for policy interventions that deviate from individual preferences.

Within this context, the concept of merit goods, originally introduced by Richard Musgrave (1957, 1959), may be helpful - for merit good arguments posit preferences that deviate from individual preferences for some specific good: Musgrave 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). For instance, 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 (Andel 1984). To see this, consider the case of health care: many governments highly subsidize health insurance, particularly for older people, to improve public health beyond what people would purchase by themselves. A policy based on purely economic rationale would contend itself with correcting informational asymmetries so as to equate marginal cost and benefit of health care.[1]

Against this background, it might be argued that merit good arguments offer such justifications in the case of sustainability as well. Indeed, Rogall (2013: 558) claims that merit goods pose an essential conceptual foundation for sustainability economics. However, the literature reveals only very few specific links. For instance, there has been made the case for organic food as a merit good (Mann 2003) on the grounds that market

[1]Certainly, health care policies also reveal striking national differences: within the US political culture, health insurance seems to be considered much closer to an ordinary economic good that needs to be individually earned, as compared to much of Europe where access to health insurance figures rather as a societal value than a consumptive good.

demand does not yet match the reflective or informed preferences for organic food. But mostly, merit good arguments are only made implicitly, without actual reference to the conceptual discussion, such as in the case of education for "mindfulness", which is argued to "contribute both to more sustainable ways of life and to greater well-being" (Ericson et al. 2014: 73). On a general level, then, this paper aims to assess what specific contributions merit good arguments may offer for discussions of sustainability.

To this end, the paper reviews two fundamental questions that are well known from the public finance literature (Musgrave 1987, Goodin 1989, Dasgupta 2005): 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? These questions obviously pertain to discussions of sustainability as well. The first question on conceptions of well-being 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 on the deviation from individual preferences directly relates to government interventions in the name of sustainability. With regard to environmental issues there is a growing literature that applies insights from behavioural economics to influence individual behaviour such as green nudges with default renewable electricity contracts (e.g. Sunstein and Reisch 2013, Croson and Treich 2014). For instance, the British "nudge unit", a social purpose company partly owned by the UK government, aims to "to encourage people to make better choices for themselves and society" in a variety of fields, such as health and safety.[2] However, such interventions in the name of Libertarian Paternalism (Sunstein and Thaler 2003) are controversial. Interestingly, within this debate critics of even soft interventions ("nudging") stand against proponents of much stronger policy interventions to foster sustainability (e.g. Schnellenbach 2016,

[2] http://www.behaviouralinsights.co.uk

Dobson 2014). This discussion on the justification of paternalist policies and the merit good literature can inform the general debate on sustainability policy and individual choice.

In what follows, we (i) clarify the definition of merit goods by considering the model for merit good arguments in Besley (1988), (ii) we 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 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 3 examines conceptions of well-being behind merit good arguments and associated justifications for paternalist policies. Section 4 discusses the challenges and opportunities of merit goods in an intergenerational framework and sustainability discussions. Section 5 concludes the discussion.

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 and transfers. Therefore, the examples do not serve particularly well to explicate the specific argument for government intervention 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 broadly as:

Definition 1 (Merit good)
A *merit good* is a good for which a social planner prescribes preferences different from actual 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, we will use the model for merit goods in Besley (1988) for clarification. For this, consider an economy with two goods and at least one representative 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, which is increasing, strictly quasi-concave and twice continuously differentiable for both goods:

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

In welfare economics, this utility function is usually 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 we call, the *merit utility function* of individual i, based on merit preferences from a social planner for good y_i.

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

Also, Besley further specifies the merit good function with $\mu_i(y_i) = \theta_i y_i$ as $V_i(x_i, y_i) := U_i(x_i, \theta_i y_i)$. 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. Consider two allocations (x_i^a, y_i^a) and (x_i^b, y_i^b) with $(x_i^a \neq x_i^b)$ and $(y_i^a \neq y_i^b)$. That is, it can be that the individual prefers an allocation over the other, i.e. $U_i(x_i^a, y_i^a) < U_i(x_i^b, y_i^b)$, yet that this choice would actually decrease merit utility, i.e. $V_i(x_i^a, y_i^a) > V_i(x_i^b, y_i^b)$. The scaling model allows to easily distinguish merit goods with $\theta_i > 1$ from demerit goods with $\theta_i < 1$.[3]

As the social planner's evaluation deviates from individual choice, that is $V_i \neq U_i$, this might lead the social planner to correct the individuals' choices 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 individual optimal merit good taxes $t_i = (1 - \theta_i)/\theta_i$ (Besley 1988: 376, Equation 3.8). It leads to the obvious result that (de-)merit goods

[3] 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 in this paper Besley's model is fine.

$\theta_i > 1$ ($\theta_i < 1$) should be subsidized (taxed). Apart from individual merit good taxes, Besley also considers the case where the state can implement only one merit good tax for all individuals. A reason for this one-size-fits-all merit good tax could be constraints on the enforceability of individual merit good taxes or lack of information on individual preferences. (Besley 1988: 378).[4]

In sum, this model allows to disentangle externalities, redistribution and merit goods. Merit good arguments concern the form of merit utility V_i and $\mu_i(y_i)$, i.e. the deviation from individual preferences for some good. For example, this could justify 'merit taxes' on cigarettes to improve peoples' health. Externalities concern external effects with given utility functions U_i that originate, for example, from the consumption of good y_i on a different economic agent. 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 could concern inequality between different individual utility levels U_i. For example, there could be transfers between individuals with high or low levels of utility. Thus, this model clarifies the specific economic concepts behind merit good arguments and differentiates it clearly from redistribution and externalities.

Note that this deviation from individual preferences sacrifices Pareto-efficiency and severely limits the applicability of welfare economics (e.g. Cowen 1993). This is in line with the discussion on many different theories of justice which violate Pareto-efficiency. For example, Kaplow (2007) shows the conflict between Pareto efficiency and ethical theories which are not based on individual preferences, but rather use the means to the fulfilment of individual preferences in their assessment of allocations (such as the capability approach by Sen (1980) and primary goods in Rawls (1971). Similarly, there is a discussion on the conflict of efficiency criteria based on different ethical theories (e.g. Kaplow and Shavell 2001, Brun and Tungodden 2004, Fleurbaey 2007). This shows how unfamiliar the idea of merit goods is to standard welfare economics.

With the background of this analytical framework, we now discuss some questions

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

on the ethical justification of taxes and subsidies on merit goods in the next section.

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:

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.[5]

Following this definition paternalism raises two questions regarding merit goods. 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? This question is addressed in Section 3.1. Second, why are government policies that are not based on individual preferences justified? This issue will be explored further in Section 3.2.

3.1 Conceptions of well being (or merit for what?)

In a first step, we consider two common conceptions of well-being regarding the utility function U_i which do <u>not</u> lend themselves to merit good arguments (as in Hausman and McPherson 2006: 118). This first step is to quickly restate the most common conceptions of well-being used to defend the inviolability of individual preferences in the name of consumer sovereignty. The first is a *hedonist conception* that says that if an individual chooses an allocation over another $U_i(x_i^a, y_i^a) < U_i(x_i^b, y_i^b)$, then this choice increases the happiness experienced by the individual, i.e. the latter allocation leads to a mental state

[5]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).

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 $U_i(x_i^a, y_i^a) < U_i(x_i^b, y_i^b)$, 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 choice in either increasing their happiness or the satisfaction of their preferences. So, both conceptions are not compatible with merit good arguments.

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 a '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, we will follow his discussion with small variations and highlight connections to Musgrave's writings.

The first conception that we 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. $U_i(x_i^a, y_i^a) < U_i(x_i^b, y_i^b)$, 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. $V_i(x_i^a, y_i^a) > V_i(x_i^b, y_i^b)$.[6]

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 the central arguments 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. $V_i(x_i^a, y_i^a) > V_i(x_i^b, y_i^b)$, 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.

In the following, we lay out how the conceptions of *'informed' preference satisfaction* and *perfectionism* translate into justifications for merit good interventions.

3.2 Justifications for merit goods

While Goodin generally speaks of a 'utilitarian' conception, 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

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

a democratic society, where an informed group is justified in imposing its decision on others" (Musgrave 1959: 14).

Two other justifications[7] 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 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

[7]Goodin considers two further justifications that relate to a utilitarian conception: one based on paternalist altruism, that he calls 'externalities', and a second based on minority moral judgements, that he calls 'high moralism'. We do not discuss these here as they do not provide justifications pertinent to this discussion.

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). In the words of Kirchgässner (2015: 22): “Paternalistic political measures can be understood as collective self-bindings via the political process”.

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 allows to incorporate concerns for sustainability (or intra- and intergenerational justice) into a democratic setting without the charge of unjustified paternalism. 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). However, merit good arguments in a narrow sense do not concern any normative objective, but rather concern the conception of well-being of individuals.

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’). In this way, all justifications relate to the definition of paternalism: they clarify what makes people better off and set standards for achieving consent from affected individuals. Furthermore, all of these justifications rely on some form of democratic accountability or directly observable acts to establish this consent. This limits the applicability of these justifications as these require some form of political arena where such consent can be achieved. For example, the justification for global merit good taxes suffers from the lack of a global (democratic) political arena. [8]

[8] These justifications allow the deviation from individual preferences in general, they do not set specific limits on how large the deviation and in turn how high merit good taxes may be. This means that these justifications do not directly prohibit either individual merit good taxes or one merit good tax for all individuals as discussed in Section 2.

4 Challenges and opportunities for sustainability discussions

In Section 3, we saw that merit good arguments are based on specific conceptions of well-being with specific justifications all of which rely on some form of consent (be it through presently or retrospectively directly observable acts or democratic political acts). In the following, we consider three general points regarding merit good arguments and sustainability: Merit goods and sustainability as justice, the challenges of future merit goods, and the relationship between green nudges and merit good arguments.

Merit goods and sustainability as justice

The first point relates to the role of merit good arguments can play in theories of justice for sustainability. When sustainability is defined in terms of (intergenerational) justice (e.g. Toman 1994, Dobson 1998, Baumgärtner and Quaas 2010) this brings up many different aspects such as conceptions of well-being and redistributive concerns. To see whether merit good arguments can bring something to this discussion it is useful to look at three minimum 'bases of distinction' on which theories of justice differ (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). As discussed in Section 2 merit goods are not an instance of redistribution which means they do not pertain to (ii) aggregation rules. Also they do not pertain to (iii) the priority of some aspect of individual advantage as these consider specific normative goals that are not included in our conceptions of well-being such as civil liberties. However, conceptions of well-being and, therefore, merit goods are relevant for (i) the metric of individual advantage in a theory of justice.

If sustainability is defined on the basis of more complex conceptions of well-being such as informed preferences or perfectionism, then this brings up merit good arguments as discussed in Section 3. In this case, the analytical results on merit good arguments can provide an opportunity for discussions on how to achieve sustainability by steering individual behaviour – beyond interventions that only aim at internalizing externalities such as Pigouvian taxes. For example, if sustainability were based on some perfectionist conception of well-being where health is defined as a value in itself, then merit good arguments could be used to discuss sin taxes on sugar and excessive meat consumption or subsidies for organic farming. Similarly, if a close relationship between humans and nature is seen as a value in itself, then recreational activities in natural surroundings could be subsidized via a merit subsidies.[9] Such interventions against actual individual preferences on the basis of sustainability as justice still would have to rely on one of the justifications for consent from affected individuals as discussed above.

A related point is whether merit good arguments are relevant for the discussion on "sustainable" social welfare functions. Such social welfare functions are widely applied in practical sustainability discussions such as the calculation of the benefits and costs of climate change (Stern 2007). To relate 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. Assume the following intergenerational *social welfare function*:

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

This definition of social welfare allows many different aggregation rules based on the metric of individual utility such as maximin, (un-)discounted utilitarianism and many more. There exists a broad literature on how intergenerational justice can be incorporated in social welfare functions, what distribution between generation results when applying these functions in different economic context and what axiomatic basis they share (e.g. Asheim 1988, Chichilnisky 1996, Alvarez-Cuadrado and Van Long 2009). For this discussion merit good arguments are irrelevant when they can be represented by a measure of individual advantage V_i with the technical properties as U_i. That is, these technical discussions leave room for different interpretations of metrics of individual ad-

[9]In a somewhat similar way, one could conceive of strong sustainability as based on the intrinsic value of nature – the resulting "perfectionist" conception of sustainability would deviate not only from individual preference satisfaction but also from the anthropocentric perspective on welfare more generally.

vantage. A *merit social welfare function*: $W(V_1, V_2)$ where merit utility is taken as the metric of individual advantage is not necessarily functionally different. Still, this often assumes comparability between V_1 and V_2 which is a controversial assumption just as in the case of comparability between U_1 and U_2. One could argue, however, that merit good arguments still allow this assumption as the paternalist streak in merit goods aims to equalize people's preferences. This possibility to bring merit good arguments into common social welfare functions follows from the difference between the focus of merit goods on individuals and distributive concerns as discussed in Section 2. That is, merit good arguments do not help the challenge on how to formulate a social welfare function that best captures the notion of intergenerational justice.

The challenges of future merit goods

A general question that arises for sustainability and merit goods is: How do the justifications of merit goods in Section 3 fare with respect to future merit goods? The usual formulation of intergenerational social welfare functions take the form as in Equation (3). One common criticism of such intergenerational social welfare functions is that they are based on the assumption that future preferences are known, here U_2, 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 for granted. The crucial problem with the formulation of a merit social welfare function is that the deviation from future individual preferences requires further information to be justified. That is, all of the justifications in Section 3 rely on directly or retrospectively observable acts or some form of democratic accountability and information on outcomes of political processes (as in Goodin 1989: 251). When dealing with future generations all of these justifications run into problems: The present generation cannot directly or retrospectively observe future acts to justify merit goods. Also there is no democratic political arena where future generations can voice their political preferences. Consequently, individuals' consent cannot be established in any meaningful way ex-ante. For example, the deviation from individual preferences in the perfectionist conception of well-being relies on the justification via 'political preferences'. Thus, using an index as V_2 that is different from the actual preferences for the second generation presupposes knowledge, not only of future preferences, but also on the outcome of democratic processes. Another question is how to incorporate future merit utility in a merit social welfare function 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 a merit social welfare function. A related challenge lies in the question of how merit utility can be 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. In sum these higher informational requirements challenge the applicability of future merit goods for the sustainability discussion.

Libertarian Paternalism, green nudges and merit good arguments

The current debate about nudging can also be considered in two objectives of policy interventions: interventions may intend to make people better-off compared to their actual choices (e.g. satisfying their long-term preferences) or may aim at achieving some distributive goal. The first issue gives rise to classic debates around merit goods and paternalism while the second is not covered by the merit goods concept (see Section 2). As an example one could consider green nudges as a means of fostering environmental goals such as climate protection, where electricity customers could be presented with a renewable-based contract as a "green default" (Sunstein and Reisch 2013, Schubert 2017).

The first point directly links the debate about Libertarian Paternalism (Thaler and Sunstein 2008) to the longstanding discussion on merit good interventions. Nudges have been proposed as softer forms of paternalism in that they only modify the choice architecture individuals face but still leave the possibility of deciding against the option suggested by the choice architect. The prime example are mechanisms designed to help individuals overcoming self-commitment problems, such as the short-term inclination to save too little for retirement. Other policy interventions more explicitly aim at steering consumption patterns so as to improve specific measures of individual well-being. For example, the UK government recently announced a "soft drink levy" that would come into force in 2018 in order to reduce obesity rates.[10] Another example would be green nudges to foster the consumption of organic foods, if these are justified with merit good arguments. Such interventions might rely on the informed preferences justifications laid out in Section 3, which build on the retrospective consent of concerned individuals or through directly observable acts, for instance expressed through the demand for health products and diet companions.

The second issue pertains to redistribution via paternalist policies associated with nudges. For instance, as a means of fostering environmental goals such as climate protection, electricity customers could be presented with a renewable-based contract as a "green default" (Sunstein and Reisch 2013). Note that these redistributive nudges are not in line with the discussion on merit goods above, as they do not aim at making people better-off compared to their actual choices. Still, there have been raised critiques against such redistributive nudges from two opposite sides. One the one hand, there are warnings of a slippery slope from soft towards strong paternalism, because even purportedly small modifications of choice architecture may not be sufficiently covered by either the informed preference or the implicit political choice argument: for example, the above green default has been criticized for being "quite openly advertised as a mechanism to circumvent political opposition in order to reach a goal that is externally imposed, rather than deduced from individual political preferences" (Schnellenbach 2016: 150). On the other hand, nudges are scolded for eschewing normative debate and this is especially relevant for sustainability policy: As sustainability inherently revolves around complex ethical judgements, the idea of facilitating sustainability via imperceptible adaptations of individual behavior amounts to a category mistake (Dobson 2014). So, while it may be possible to achieve redistribution from the present generation to future generations via green defaults such as renewable-based contracts, this approach fails on the account that it inhibits discussion on normative trade-offs on sustainability. That is, without environmental citizenship, without individuals explicitly reflecting on their behavior and

[10]https://www.gov.uk/government/publications/budget-2016-documents/budget-2016

their attitudes in the context of intergenerational justice, there will be no attainment of sustainability as a normative concept that entails some notion of individual liberty (Dobson 2007).[11] In sum, redistributive green nudges might look tempting to ensure the attainment of redistribution to future generations, but might not be sufficient to resolve the complex normative questions around sustainability. Still, the issue of redistribution lies outside the common discussion on merit good arguments.

In the best of circumstances, however, green nudges based on merit good arguments might support both preference satisfaction and intergenerational redistribution. This would be the case when there is a connection between a merit good and negative intergenerational externalities: then a green nudge on the merit good would improve the well-being of future generations. For instance, merit taxes on sugar and red meat could also reduce potential externalities from agricultural production such as biodiversity loss, land degradation and the like (e.g. Tukker et al. 2011, Vinnari and Tapio 2012). In this sense, merit good arguments provide the opportunity to study the interaction between different arguments for government intervention on one particular good in the context of sustainability policy: namely Pigouvian taxes to increase Pareto-efficiency with given individual preferences and merit taxes to make people better-off compared to their actual choices.

5 Conclusion

In this article, we discussed the origin of the concept of merit goods, its current application in economic analysis and its potential for sustainability discussions. While merit good arguments are invoked in many contexts, we saw that their definition and ethical justification remains quite vague. For this, we discussed different conceptions

[11]However, there may also be nudges that stimulate reflective thinking – so-called type 2 nudges, as compared to type 1 nudges that purely work through intuitive and subconscious channels (Hansen and Jespersen 2013). Type 2 nudges aim at encouraging rather than circumventing conscious reflection (for instance, energy bills that offer social comparisons of consumption levels). Following Dobson's line of argument, such type 2 nudges would be clearly preferable.

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 or consent – which is similar to what Musgrave argued initially. We also saw that merit good arguments, that deviate from individual preferences, lend themselves to analyze the consequences for policy intervention: While a Pigouvian tax achieves Pareto efficiency for given individual preferences, a merit good (e.g., through public health as a 'perfectionist value') may justify a higher tax level than the one required for Pareto-efficiency. That is, merit good arguments sacrifice Pareto-efficiency in the standard welfare economic sense.

As to the question of what merit good arguments bring to sustainability discussions, we focused on three general points:

First, we saw that merit good arguments concern conceptions of individual wellbeing and are, therefore, clearly different from distributional arguments in favour of intergenerational equity. This means that merit good arguments do not contribute a novel argument in favour of making future generations better off by conserving environmental assets and transferring more resources to the future. Also, this means that merit good arguments are not directly relevant to the discussion on which social welfare function best captures intergenerational justice.

Second, by discussing the general hurdles to justify a deviation from individual preferences in an *intragenerational* context, it became clear that merit good arguments can only serve a narrow purpose under very special circumstances. When merit goods are defined for future generations these hurdles become even greater as justifications require additional information – for example, when future 'political preferences' are required to justify the deviation from individual preferences via merit good taxes. These hurdles become so great that it seems very difficult to exploit the concept of merit goods fruitfully for future generations.

Third, although merit good arguments cannot broadly enrich the discussion of intergenerational environmental problems, they can shed some light on the ethical challenges raised by green nudges. If green nudges directly aim at redistribution to future generations, as in the case of climate mitigation through default renewable energy contracts, merit goods cannot deliver an explicit ethical justification as this would require the aim to make affected present individuals better-off and getting some form of consent from them. Consequently, such green nudges may well succeed in redistributing from present to future generations, though only by dodging public debate, clouding the discussion on normative trade-offs and interfering with individual freedom. However, there might be cases where merit good arguments justify policy intervention on behalf of individuals' reflective long-term preferences (e.g., "sin taxes" on red meat or sugar) and where the resulting shift in consumption patterns possesses the welcome side-benefit of reducing intergenerational environmental externalities and making future generations better-off in the process.

In sum, what contribution can merit good arguments offer in sustainability discussions? On the one hand, a general reflection on the justifications for merit goods is helpful to disentangle specific motivations for policy interventions and to analyze sustainability problems beyond the standard setting of welfare economics. This also connects to the ongoing debate on the normative foundation of economic analysis (e.g. Sen 1977, Hausman and McPherson 2006, Sandel 2013) and the potential gains in a systematic discussion on conceptions of well-being between ethics and economics. On the other hand, however, it should be clear that merit good arguments do not offer a blank check to correct individual preferences in the name of sustainability. Rather, whenever merit goods are proposed for a new sustainability policy which deviates from individual preferences, then there must follow a thorough assessment of the legitimacy of this intervention. As sustainability is (over-)laden with normative assumptions, such policies require careful economic and ethical examination.

Acknowledgements

We are grateful to Dave Abson, Stefan Baumgärtner, John-Oliver Engler, Joachim Fünfgelt, Sergio Mukherjee and Michael Schefczyk for critical and constructive discussion and the German Federal Ministry of Education and Research for financial support under grant 01UN1011A. All remaining errors are our own.

References

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.

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.

Bullock, C.H. and Collier, M. (2011). When the public good conflicts with an apparent preference for unsustainable behaviour. *Ecological Economics*, 70(5):971-977.

Cowen, T. (1993). The Scope and Limits of Preference Sovereignty. *Economics and Philosophy*, 9:253–269.

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

Croson, R. and Treich, M. (2014). Behavioral Environmental Economics: Promises and Challenges. *Environmental and Resource Economics*, 58(3):335–351.

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

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

Dobson, A. (2007). Environmental Citizenship: Towards Sustainable Development. *Sustainable Development*, 15:276–285.

Dobson, A. (2014). *Sustainability Citizenship.* Green House Publications, http://www.greenhousethinktank.org/page.php?pageid=publications.

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/`]

Ericson, T., Kjonstad, B. and Barstad, A. (2014). Mindfulness and Sustainability. *Ecological Economics*, 104(3):73–79.

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.

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

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, pp. 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, pp. 357–359.

Hansen, P.G. and Jespersen, A.M. (2013). Nudge and the Manipulation of Choice. *European Journal of Risk Regulation*, 4(1):3–28.

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.

Kirchgaessner, G. (2015). Soft Paternalism, Merit Goods and Normative Individualism. *European Journal of Law and Economics*, forthcoming.

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, pp. 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.

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

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

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

Rogall, H. 2013. *Volkswirtschaftslehre fŸ̈r Sozialwissenschaften.* Springer, Wiesbaden.

Sandel, M. J. (2013). Market Reasoning as Moral Reasoning: Why Economists Should Re-engage with Political Philosophy. *Journal of Economic Perspectives*, 27(4):121–40.

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. (2012). Nudges and norms: on the political economy of soft paternalism. *European Journal of Political Economy*, 28(2):266–277.

Schnellenbach, J. (2016). A constitutional economics perspective on soft paternalism. *Kyklos*, 69(1):135-156.

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.

Schubert, A. (2017). Green nudges: Do they work? Are they ethical?. *Ecological Economics*, 132:329–342.

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, pp. 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, pp. 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.

Sunstein, C.R. and Reisch, L. (2013). Green Defaults. *Kyklos*, 66:398–402.

Sunstein, C.R. and Thaler, R.H. (2003). Libertarian Paternalism is Not An Oxymoron. *University of Chicago Law Review*, 70:1159–1202.

Toman, M. (1994). Economics and "Sustainability": Balancing Trade-Offs and Imperatives. *Land Economics*, 70(4):399–413.

Thaler, R.H. and Sunstein, C.R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Penguin, London.

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

Tukker, M., Goldbohm, R.A., de Koning, A., Verheijden, M., Kleijn, R., Wolf, O., Perez-Dominguez, I. and J.M. Rueda-Cantuche (2011). Environmental impacts of changes to healthier diets in Europe. *Ecological Economics*, 70:1776–1788.

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/`]

Vinnari, M. and Tapio, P. (2012). Sustainability of diets: From concepts to governance. *Ecological Economics*, 74:46–54.