The economics of climate change with endogenous preferences*

Linus Mattauch[†] Cameron Hepburn[‡] Fiona Spuler[§]
Nicholas Stern[¶]

June 15, 2021

*We thank Inge van den Bijgaart and Jan Siegmeier for numerous helpful discussions and Samuel Bowles, Partha Dasgupta, Ottmar Edenhofer, Karla Hoff, Franck Lecocq, Gordon Menzies, Michael Pahle, Alexander Teytelboym and seminar audiences at CIRED, Columbia, Oxford, ETH Zurich, the World Congress of Environmental and Resource Economists 2018 and the EEA, LAGV and RES meetings 2019 for useful comments. Simona Sulikova provided excellent research assistance. Linus Mattauch's research was supported by a postdoctoral fellowship of the German Academic Exchange Service (DAAD) and the Robert Bosch Foundation. Cameron Hepburn thanks the Post Carbon Transitions programme (LDR00530) from the Oxford Martin School. Nicholas Stern acknowledges funding support from the ESRC Centre for Climate Change Economics and Policy and the Grantham Foundation for the Protection of the Environment. This manuscript is also part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 681228.

[†]Institute for New Economic Thinking at the Oxford Martin School and Environmental Change Institute, School of Geography and the Environment, University of Oxford

[‡]Smith School of Enterprise and the Environment, Institute of New Economic Thinking at the Oxford Martin School, School of Geography and the Environment, University of Oxford

[§]Institute for New Economic Thinking at the Oxford Martin School, University of Oxford

[¶]Grantham Research Institute, Centre for Climate Change Economic and Policy, and Department of Economics, London School of Economics

Title: The economics of climate change with endogenous preferences

Abstract

Avoiding unmanageable climate change implies that global greenhouse gas emissions must be reduced rapidly. Carbon prices and technological development are essential to deliver such reductions. Changes in preferences, however, are rarely considered, even though other major socioeconomic transitions – such as those from reducing smoking and drink-driving – have succeeded partly because preferences have changed. This article examines the impact of climate policy-induced changes in consumers' preferences. We demonstrate that when policies also change preferences, they are inefficient unless such changes are accounted for. For instance, carbon taxes must be adjusted, if they crowd-in or -out social preferences, to achieve a given target. Further, when the urban built environment changes mobility preferences, the value of low-carbon infrastructure investments can be underestimated if such effects are ignored. Third, policy-induced changes in preferences for active travel and plant-based diets could increase the net benefits of the transition to zero emissions.

JEL codes: A12, D91, H23, Q54, Q58

Keywords: climate change, carbon pricing, endogenous preferences, crowding-in, transport infrastructure, health co-benefit

1 Introduction

To deliver on the temperature targets of the Paris agreement on climate change, greenhouse-gas emissions must be reduced to net zero within decades (IPCC, 2013, 2018). Policy instruments such as carbon prices should make a crucial contribution to this challenge for public policy, in combination with other instruments, such as major investments into low-carbon infrastructure, and also more gentle 'nudges'. Changes in preferences — which go beyond nudges — are, however, rarely considered as part of the strategies to reduce emissions, even though other major socioeconomic transitions have succeeded partly because preferences have changed. This article explores the implications for evaluating climate change mitigation instruments when such policies can also change preferences, intentionally or unintentionally. Within environmental economics, there is now a small literature on endogenous preferences (Perino, 2015; D'Haultfœuille et al., 2016; van den Bijgaart, 2018; Konc et al., 2021), norms (Ulph and Ulph, 2018; Dasgupta et al., 2016), beliefs (Koessler and Engel, 2019) and culture (Schumacher, 2015). Yet it does not address how to achieve socially optimal environmental policy when it changes preferences and does not apply the main insights to climate policy design

Much research in economics understandably assumes that preferences are exogenous – following the liberal dictum that "de gustibus non est disputandum" (Stigler and Becker, 1977) – for the compelling reason that individuals should have the freedom to develop their personalities in any way they choose, without government intervention. For example, when a cost-benefit analysis of a large infrastructure project is carried out or a tax rate is recommended, changes in preferences caused by such policies are not taken into account.

However, an increasingly large body of evidence within economics and other social sciences indicates that preferences can be endogenous to policy decisions (Bowles, 1998; Akerlof and Kranton, 2000; Alesina and Fuchs-Schündeln, 2007; Fehr and Hoff, 2011; Fuchs-Schündeln and Schündeln, 2015; Hoff and Stiglitz, 2016). Even our values[1] and culture[2] – upon which preferences are based (Hoff and Stiglitz, 2016) – are amenable to change by policy. This is widely understood in other social sciences (Fehr and Hoff, 2011; Hawkes et al., 2015). Within economics, endogenous preferences have been discussed in the past (Gintis, 1974; Pollak, 1978; von Weizsäcker, 1971), but the analysis of the welfare effects has only been developed more recently (Binder, 2010; Fleurbaey and Tadenuma, 2014; von Weizsäcker, 2005) and

[1]'Values' are defined by the Oxford English Dictionary as 'the principles or moral standards held by a person or social group; the generally accepted or personally held judgement of what is valuable and important in life'.

[2]'Culture' is relevantly defined as 'a way of life or social environment' and as 'the philosophy, practices, and attitudes of an institution, business, or other organization'.

not yet been applied to environmental policy specifically.

The idea that preferences can be endogenous is distinct from the large body of literature on behavioural economics and public policy (Allcott et al., 2014; Bernheim and Rangel, 2009; Farhi and Gabaix, 2020; Thaler and Sunstein, 2008): behavioural economics, which includes the study of 'nudging', concerns realistic human behaviour, given context, and given so-called 'irrational' cognitive biases. Changes in values and culture, however, are not 'irrational' because these are elements that economists tend to consider to be logical priors. But changes in values and culture can drive long-lasting changes in behaviour[3] as shown by changes in attitudes to issues such as smoking, drink-driving and recycling (Convery et al., 2007; Nyborg et al., 2016).

This article has two core purposes. First, we explore the idea that low-carbon policies could be better designed if it is recognised that preferences can be endogenous to such policies. We review evidence in three areas in which policy affects preferences: (i) the impact of carbon prices on preferences for low-carbon consumption options; (ii) the impact of transport infrastructure on mobility preferences; and (iii) the impact of policy on preferences for low-carbon diets and active travel.

Second, we develop a highly-stylised theoretical framework to capture the essence of the ideas in a relatively general way that is agnostic to the detailed cognitive and social mechanisms involved, and use the framework to elucidate a central feature of the three different examples. The model makes some heroic simplifying assumptions; preferences, values and habits are considered to be observationally equivalent at the high level of abstraction we are considering.[4] However, the heroic simplifying assumptions allow us to obtain general results on how key recommendations about policy instrument choice change if preferences are endogenous. For evaluating climate change mitigation options by public finance methods, the model may be used to analyze different mechanisms for changing the strength of preferences by policy, including taxation making consumers less or more altruistic, infrastructure affecting habits and information campaigns affecting individuals' moral principles.

With this high level of abstraction, welfare analysis – in the sense of some policies being considered to be superior to others – is possible and fairly straightforward, notwithstanding the endogeneity of preferences. This

[3]It is established that values shape human actions by influencing emotions, and that values are acquired through social interactions and learning experiences (Schwartz, 1994); see Akerlof (2017) and Roos (2018) for applications in economics and Corner et al. (2014) for climate change.

[4]Related contributions enter individual psychological or social mechanisms by which policies shape preferences in their models (van den Bijgaart, 2018; Ulph and Ulph, 2018). Values, especially in the sense of Schwartz (1994) are sometimes seen as more fundamental than preferences, yet this is not what we are concerned with in this article.

is true if we assume either that utility is cardinal, or if we posit some form of 'meta-preference'. We can also answer questions such as whether a change in preferences for low-carbon consumption might contribute to meeting exogenously given environmental constraints (such as national emission reduction targets) if regulation is inadequate.

The model permits us to answer the questions how appropriate policy is different when it is recognised that it can shift preferences in favor of low-carbon options. Specifically, we demonstrate the following results. First, if carbon pricing changes consumers' preferences, not merely relative prices, then the policy will be inefficient unless it is adjusted to account for this endogeneity. For instance, when a carbon tax has an impact on preferences, the preferred tax rate will depend on whether there is crowding-in or crowding-out. We show that there is a unique level of crowding-in at which the tax with endogenous preferences coincides with the conventional tax with fixed optimal preferences; otherwise, the two tax rates are different. Second, we show that preference endogeneity can change public investment decisions: if low-carbon infrastructure itself increased the desirability of consuming low-carbon goods, the marginal value of investing into such infrastructure is higher. Third, when preferences for active travel and low-carbon diets are endogenous to policy intervention, such policies are more desirable given the substantial health benefits (in addition to the low-carbon benefits) from such changes in preferences. Overall, our framework allows us to trace the consequences of preference endogeneity for environmental policy.[5]

We wish to be explicitly clear that we are not advocating for interventions to undermine people's freedom to develop their own objectives or to make their own choices. Rather, our framework suggests that we should account for the impact of policy on preferences, directly or indirectly, and whether we like such impacts or not. Failure to account for these effects leads to suboptimal policy outcome. Public dialogue and discussion can inform the evolution of societal preferences, without violating freedoms and resolve conflicts between them. For instance, societal attitudes to and preferences for tobacco have evolved, reducing the freedom to smoke in some public spaces, but increasing the freedom of non-smokers to enjoy such spaces without increased risk of cancer. In the cases of smoking and drink-driving, societies chose to guide the processes of shifting preferences (Stuber et al., 2008; Levy et al., 2012; Mons et al., 2013; Watling and Armstrong, 2015) and

[5]Our approach is indeed reminiscent of Stigler and Becker (1977) in that we enrich the arguments of utility a function, going beyond consumption of market goods. However, we are not, like Stigler and Becker (1977), interested in explaining behaviour by generalised consumption goods, but instead study the normative consequences of the fact that those arguments change with policy. This applies whether or not the change in utility as a consequence of a policy is either due to a change in preference reflecting moral values, say, or else is merely capturing a psychological reaction to changed infrastructure or to other Beckerian generalised consumption goods. These two possibilities are formally equivalent in the approach pursued here.

information and education campaigns took place, alongside price interventions and bans (Stern, 2015, Ch. 10). The scale of the challenge of reducing greenhouse gas emissions to net zero suggests that a similar combination of approaches may be desirable. Shifts in preferences for transport, energy and food choices will protect freedoms by reducing deaths and impacts from climate change.

Our contribution builds on several important articles. Bowles and Hwang (2008) examines optimal public good provision when policy changes preferences, and Frey and Stutzer (2008) and Bowles and Polania-Reyes (2012) discuss crowding-in and -out of intrinsic motivation by policy measures (see also the more general Bénabou and Tirole (2003, 2006) on intrinsic and extrinsic motivation). Farhi and Gabaix (2020) also study modified Pigouvian tax rules, but due to behavioural biases, not changes in preferences. Bisin and Verdier (2001) study the intergenerational transmission of preference traits. Bezin and Ponthière (2019) study the implications of such a transmission for the dynamics of different moral behaviors in a Tragedy of Commons situation. Schumacher (2015) considers the relationship between culture and environmental quality. van den Bijgaart (2018) studies the optimal transition policy when habits are affected by past consumption decisions (Pollak, 1970; Ryder and Heal, 1973) and consumers fail to internalise shifts in habits. Ulph and Ulph (2018) find that taxes can be welfare-reducing when individuals adjust their consumption to conform to the norms of a group to which they wish to belong. Dasgupta et al. (2016) examine environmental policy given consumption norms and the social context of consumption. Daube and Ulph (2016) examines how preferences can help achieve objectives in situations of inadequate regulation. Jacobsen et al. (2012) develops a theory of voluntary public good provision when households are motivated to offset the environmentally harmful behaviour of others. Bezin (2015) explains private provision of environmental public goods based on a desire to socialize others into having environmental preferences, showing how this can account for differences between environmental attitudes and behaviour.Brennan (2006) and Perino (2015) examine the effectiveness of green preferences as a function of the policy setting. In addition, Bezin (2019) studies the relationship between green preference formation and the direction of technological change, highlighting the role of non-monetary policy instruments to avoid carbon lock-in. We differ from all these prior contributions by characterising how the possibility of preference changes by policy affects recommendations on environmental taxation. Unlike almost all of those contributions, we focus on the question of optimality rather than pursuing a positive approach to the interaction of preferences and policies.

The remainder of this article is structured as follows: The next section describes the evidence that preferences are endogenous to policies and institutions, it then relates this thesis to specific examples relevant to the transition to the low-carbon economy. Section 3 constructs model of endogenous preferences and carbon pricing. Section 4 extends this model to the cases of preference formation by (transport) infrastructure, as well as of the health benefits from low-carbon diets and active travel and briefly explores normative and policy implications. Section 5 concludes.

2 Endogenous preferences: the evidence

There is a wealth of evidence in psychology and sociology that the underpinnings of human choice – preferences, beliefs and decision-making processes (DellaVigna, 2009) – are culturally formed (Bowles, 1998; Hoff and Stiglitz, 2016). Preferences appear to be shaped by cultural transmission, and relate to our social identities (Akerlof and Kranton, 2000), and our worldviews and narratives (Hoff and Stiglitz, 2016).[6]

In contrast, behavioural economics and behavioural public policy study behaviour, rather than underlying preferences, cultures and worldviews. The behavioural economics and policy literature has shown that behaviour is context-dependent and may appear to be frequently "irrational". These insights are behind various important policy interventions, such as "nudges" (Allcott et al., 2014; Bernheim and Rangel, 2009; Kahneman, 2011; Thaler and Sunstein, 2008). Much valuable progress has been made in these fields in recent decades. Here, however, we explore the role of changing the preferences underlying behaviour, rather than nudging the behaviour per se (see also Hoff and Stiglitz, 2016).

Within the literature on endogenous preferences, Bowles (1998) provides a comprehensive review of evidence from biology, psychology and sociology on the formation of preferences in market economies. He argues that the primary channel for the development of preferences is cultural transmission, and finds that many economic incentives tend to negatively affect intrinsic motivation – behaviour is no longer driven by internal rewards (Bowles, 2008). The assumption of fixed preferences limits the "explanatory power, policy relevance, and ethical coherence" of economic analysis (p. 75). Two surveys corroborate these conclusions. First, Fehr and Hoff (2011) find that preferences are prone to direct social influences; social institutions stimulate certain parts of people's identities through framing and anchoring effects.[7] They conclude that "[E]xogenous preferences is but a special and

[6]See Villacorta et al. (2003) for pro-environmental behaviour in particular. Further, Voors et al. (2012) and Cavatorta and Groom (2020) find that exposure to violent conflict changes preferences; O'Hara and Stagl (2002) and Russell and Zepeda (2008) consider how preferences may change through participation in community-supported agriculture.

[7]For example, Tompson et al. (2015) find, based on neurological evidence, that "personally and culturally tailored messages" (p. 58) lead to greater neural activation that causes greater subsequent behaviour change. Thus, it is argued that institutions can shape preferences by rendering particular identities, values and norms, more salient.

not very plausible example [...] among the possible set of assumptions about preferences that can be employed in explaining economic outcomes" (p. F409). Second, Hoff and Stiglitz (2016) concludes that preferences are formed through the social context and the use of cultural mental models to process information. Examples providing evidence for these conclusions are Deckers et al. (2020), who find a relationship between the prosociality of the mother and the prosociality of the child when at primary school; Tompson et al. (2015), who find differences between culturally salient messages for Western and non-Western societies; Algan and Cahuc (2010) on parent-child transmission of trust and economic preferences, and Malmendier and Nagel (2011) who show preferences for financial risk depend on experienced stock market returns.

Given this evidence, in this paper we do not unpick the psychological and social processes through which policies change preferences.[8] We simply assume that policies can change preferences, and focus on the implications for the economics of climate change mitigation. In terms of simple microeconomics, we can represent a change in preferences by shifting utility curves rather than by rotating the budget line caused by a change in relative prices (see Figure 1). The consequence of taking the evidence on changing preferences seriously is that decarbonisation policy should account for potential shifts in preferences, and not merely focus on changing relative prices.

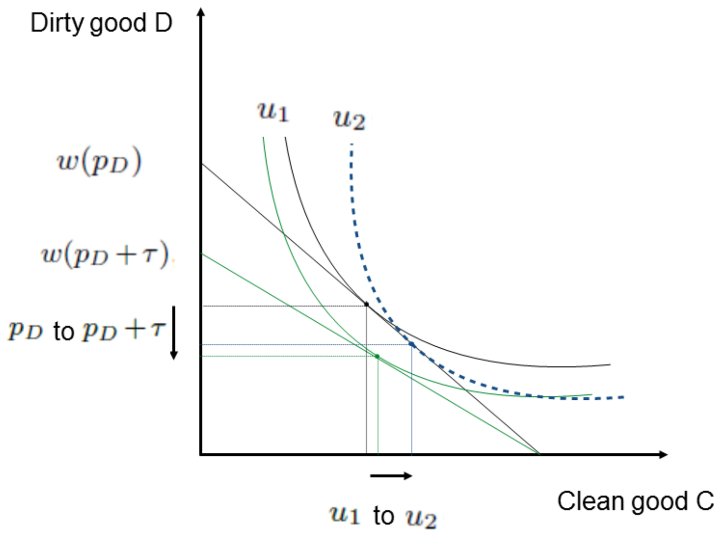
2.1 Three relevant examples

We now discuss three examples of endogenous preferences that are relevant to climate change mitigation policy, focussing on carbon-intensive and low-carbon consumption options in food and transport. This approach can be applied more broadly, however, for instance for energy efficiency, i. e. creating preferences that increase demand for energy-efficient options (Costa and Gerard, 2018; Hahn et al., 2016). In Sections 3 and 4 we formally explore the consequences for policy design.

Pricing shapes low-carbon preferences: Several examples suggest that environmental pricing and subsidies can crowd-out or crowd-in environmental values, depending on the context – that is, the incentives increase or decrease intrinsic motivation to protect the environment. For instance, building on Perino et al. (2014), Lanz et al. (2018) find that a carbon price of £19/tCO_2 on food crowds-out intrinsic motivation to the extent that compensating for this effect requires the carbon price to rise by as much as £48/tCO_2. Frey and Oberholzer-Gee (1997) find that willingness to accept

[8] A literature in cognitive psychology finds that preferences are *not* stored in memory and retrieved, but are constructed when elicited (Lichtenstein and Slovic, 2006; Weber and Johnson, 2009).

Figure 1: Simple microeconomics of a shift towards less polluting consumption.



Preferences are commonly assumed to be fixed, so that only changes in (relative) prices affect the allocation of consumption between dirty and clean goods. Consider utility indifference curve u_1 at income w. An tax τ on the dirty good, D, increasing the price from p_D to $p_D + \tau$, rotates the budget constraint from $w(p_D)$ to $w(p_D + \tau)$, reducing dirty consumption and lowering utility, based on the same preferences represented by u_1. In contrast, a shift in preferences can be represented as a move in utility function from u_1 to u_2, which leads to lower consumption of D without any change in prices or budget.

a nuclear waste site fell dramatically when monetary compensation was offered. Bowles and Polania-Reyes (2012) observes incentives may "...affect the process by which people learn new preferences" (p. 368). Lanz et al. (2018) and Mattauch and Hepburn (2016) conjecture that when pricing crowds out environmental preferences, the appropriate carbon price would need to be set higher than if preferences were fixed. Below we prove this conjecture and characterise first- and second-best policy options (Section 3). In contrast, D'Haultfœuille et al. (2016) show that a "feebate" for buying cars (i.e. a financial reward for low-emitting and a penalty for high-emitting cars) led to crowding-in effects beyond the price effects of the policy. Convery et al. (2007) find that the Irish plastic bag tax crowds-in environmental values. In British Columbia and Sweden, the salience of the carbon tax led to greater behaviour change than would be expected from an equivalent increase in gasoline prices (Andersson, 2019; Rivers and Schaufele, 2015), a finding compatible with crowding-in environmental values.

Urban transport infrastructure shapes mobility preferences: Weinberger and Goetzke (2010) provide evidence that preferences for car ownership are determined by the built environment individuals are used to. When people move from a city with good public transport to a car-dependent city they 'export' their mobility preferences to the new environment. They are more likely to own fewer vehicles due to learned preferences for lower levels of car ownership (Weinberger and Goetzke, 2010). Appropriate transport infrastructure is thus not only required to make low-carbon travel possible, but can also be a pre-condition for the learning of new mobility preferences. It is likely that this is mediated through peer effects (Grinblatt et al., 2008; Weinberger and Goetzke, 2010, see also Mattauch et al., 2016). The finding is indicative of the well-established fact that urban form is a key determinant of energy consumption in cities (Newman and Kenworthy, 1999; Seto et al., 2014). Severen and Van Benthem (2019) provide evidence for the related result that in the U. S. the 1979 oil crisis negatively influenced the propensity of those coming of driving age in that year to drive to work in 2000. Below we formally show that, low-carbon infrastructure shaping mobility preferences increases the optimal level of such investment.

Public health policy shapes dietary and transport mode preferences: Hawkes et al. (2015) describe the ways in which public health policies can help people "learn healthy food preferences". It is a societal choice as to which dietary habits should be fostered, consistent with robust evidence that diets are strongly influenced by cultural factors (Rozin and Vollmecke, 1986; Birch, 1999). This is relevant to climate policy: Springmann et al. (2016) find that, by 2050, a shift toward more plant-based diets in line with standard dietary guidelines could reduce food-related greenhouse gas emissions by 29–70 % and decrease global mortality by 6–10 %.

Similarly, policies forming preferences for active travel modes have a health and environmental benefit. Increasing active travel in urban transport (such as walking and cycling) reduces obesity-related diseases, dementia and depression, and can also reduce local and global emissions Woodcock et al. (2009).[9] Dietary and travel choices might be understood as arising from internally conflicting preferences (Loewenstein and Prelec, 1992;

[9]Numerous studies confirm the extensive health benefits from active travel increases. Woodcock et al. (2009) explore a scenario with greater active travel (slight increase in the distance walked and double increase in distance cycled), finding that the health gains from physical activity are greater than those from reduced air pollution in both London and Delhi. Rabl and De Nazelle (2012) calculate the benefits of switching from a car to cycling for trips under 5 km to be about 1300 Euro per person per year. Götschi et al. (2015) compare the health benefits that would accrue to the English population if they increased the proportion of active travel to the levels of Switzerland, California and the Netherlands and Mueller et al. (2017) systematically estimate what the health

Loewenstein, 1996)) between unhealthy short-term desires and the long-term preference for good health. However, changes in preferences are also relevant, and the evidence suggests that they influence choice of travel mode (Hopkins, 2016; Hunecke et al., 2001; Steg, 2005) and diets (Allen et al., 2000). We formally discuss these ideas below.

To the best of our knowledge, in none of the three cases illustrated above have the implications for adequate climate policy design been drawn by means of economic models, which is what we turn to next.

3 A model of a carbon price that changes preferences

We construct a simple model to explore the potential importance of shifts in consumer preferences by policy, that is, we are interested in cases in which there is an (intended or unintended) impact of any policy that changes relative prices on the utility function itself, rather than only upon the choice set (see Figure 1). We represent the currently dominant setting of a national environmental tax reform: a nation state sets a climate target (given in percentage point reduction of GHG emission, not determined by cost-benefit analysis) and then determines appropriate policy instruments. Consistently, a social planner problem in the below is a "cost-effectiveness" analysis and "first best" refers to decentralised settings reproducing such a social optimum. At the end of this section we explain how results change if this setting is instead one of "cost-benefit analysis" mode.

After introducing the model in Subsection 3.1, we first set out some basic properties (Subsection 3.2): if carbon prices are insufficient and dirty goods are cheaper, a consumer is better off if her utility functions lead her to favour dirtier goods under some regularity (Proposition 1). If policy is aiming to meet a specific climate target, however, a shift in the utility function can help compensate for inadequate carbon pricing. We then illustrate the main result of this section: when a carbon tax has an impact on preferences, the second-best tax differs from the conventional first-best tax depending on whether there is crowding-in or crowding-out (Proposition 2) – and we explore the strength of this effect. Furthermore, we demonstrate that there is a particular strength of crowding-in at which the second-best tax coincides with that first-best tax with fixed optimal preferences, but no crowding. The reason is that the crowding-in additionally helps to achieve the environmental target (Proposition 3). Finally, we discuss extensions to optimal regulation.

We represent the influence of policy choices on consumers' preferences by

changes would be for populations in seven different cities in Europe following an increase in infrastructure encouraging active transport. For a comprehensive review see also Mueller et al. (2015) and Sulikova et al. (2020) for implications for optimal urban transport policy.

augmenting the utility function with a parameter, α, that can be shifted by policy. Our aim is a simple formulation that can capture different underlying observationally-equivalent phenomena: changes in culture, values or habits. In a technical sense, we will refer to α as the consumer's 'appreciation' of the clean good.[10] An increase in the appreciation of a good that the consumer already consumes can increase utility. One can think of α as resulting from social influence, reflecting learned or inherited values or tastes. Although not our primary focus, we make the assumption that we can meaningfully think of changes in 'appreciation', and thus values, as welfare-relevant, that is, we can compare utilities for different α in a meaningful way. This is true if utility is cardinal or if consumers have preferences over their appreciation (dependent on the allocation of goods), as discussed in Subsection 4.2. In sum, the link between α and the structure of preferences is straightforward: higher α places greater weight on the clean good, C, in Equation (1), and thus re-shapes orderings and indifference curves. The relationship of α to welfare is far less straightforward, however. On the one hand, a consumer may experience greater utility from higher α from a positive self-image. On the other hand, α might rise as appreciation of the scale of the damage from pollution increases, which might the individual miserable, reducing utility. So there are plausible cases in which welfare decreases or increases as α increases.

3.1 Basic approach

Suppose a single consumer has a simple choice between two goods, one relatively clean C and one relatively dirty D, where the consumer's utility is also a function of the appreciation parameter α. First consider the social planner problem:

$$\max_{C,D,\alpha} U(C, D, \alpha) \tag{1}$$

subject to

$$p_C C + p_D D = w - \xi D \tag{2}$$

where w is income, ξ is the damage intensity of dirty consumption – effectively environmental damage reduces income – and p_C and p_D are both consumer prices and production costs, since we abstract from modelling production. We assume $U_C, U_D > 0$ and impose more structure on the interaction between goods and appreciation by positing

$$U(\alpha C, (1 - \alpha)D). \tag{3}$$

[10] The OED defines 'appreciation' as the action of 'assessing the nature or quality of something or someone; judgement, estimation.' See also Becker and Mulligan (1997), who incorporate 'appreciation' into their utility function. In their model, appreciation captures the vividness of the future, rather than a set of environmental values as here.

to ensure that relevant optimisation problems have a solution. The functional form is related to those used in the literature on status-seeking and habits, where α represents the learned or inherited preference from social influence (Abel, 1990; van den Bijgaart, 2018).

Consider a society intent on achieving a particular target (relaxed for Proposition 4 below), expressed here as a fixed amount of the dirty good, $\widetilde{D} > 0$, which in our model also leads to fixed emissions. This determines consumption of the clean good via the budget equation: $\widetilde{C} = (1/p_C)\left(w - (p_D + \xi)\widetilde{D}\right)$ and by maximising utility over the budget constraint.[11]

Now suppose the social planner can influence the level of appreciation of the clean good. For most of the below analysis, we use the specific constant-elasticity-of-substitution utility

$$U(\alpha C, (1-\alpha)D) = [(\alpha C)^\gamma + ((1-\alpha)D)^\gamma]^{\frac{1}{\gamma}} \tag{4}$$

with $0 < \gamma < 1$, meaning that clean and dirty consumption are substitutes and $\alpha \in [0,1]$ to ensure the problem has a solution.[12]

With this parametrisation, the socially optimal appreciation α^{SO} to achieve the agreed emissions target is:

$$\alpha^{SO} = (\frac{\widetilde{D}}{\widetilde{C}})^{\frac{\gamma}{\gamma-1}} / (1 + (\frac{\widetilde{D}}{\widetilde{C}})^{\frac{\gamma}{\gamma-1}}). \tag{5}$$

This equates the marginal utility loss from a reduced value of dirty consumption with the marginal utility gain from a higher value of clean consumption.

3.2 Properties of the decentralised equilibrium: optimal appreciation and first-best

Now consider a representative consumer with the same utility function as as above, but who ignores the production externality, ξD and also faces a (unit) tax τ on dirty consumption (which is fully recycled to her) – a standard set-up of studying pollution regulation. Suppose that she does not by herself

[11] This is only the case in a setting with two goods, which is the simplest possible for our research questions. In practice, consumers will optimise choice over a variety of products, and abatement possibilities, but subject to a constraint on total emissions.

[12] One might think that it is more natural to consider the parametrisation $[\alpha C^\gamma + (1-\alpha)D^\gamma]^{\frac{1}{\gamma}}$. While this indeed simplifies the process of calculating optimal appreciation for optimal environmental regulation, the problem of optimising appreciation for an environmental target then has no meaningful solution. Further, while considering $0\gamma < 0$ in the following results is possible, the case of clean and dirty goods as complements is of limited relevance for sustainability transitions.

adjust appreciation, and denote appreciation by α^M. (where M stands for "market"). Assuming the same parametrisation as above, her problem is:

$$\max_{C,D} U(\alpha^M C, (1-\alpha^M)D) \tag{6}$$

subject to

$$p_C C + (p_D + \tau)D = w + L. \tag{7}$$

L is the part of the budget of which the dependency on the tax and the damages are ignored by the consumer, $L = \tau D - \xi D$.

With variable appreciation, the following basic properties of the model can be established:

Proposition 1. *The optimal appreciation in the imperfectly regulated decentralised case ($\tau \neq \xi$) differs in general from the optimal appreciation in the socially optimal allocation, provided that $U(C) = U(\alpha C, (1-\alpha)D(C))$ is a strictly concave function in the domain given by $\alpha \in [0,1], p_c C \in [0, w]$.*

Proof. See Appendix A, including application to the above paramentrisation. □

The proposition implies that it can be advantageous to adjust preferences to the level of environmental protection: a consumer can get more total utility if she puts more value on the consumption option that is cheaper. Without the concavity assumption, it is possible that the maximum value of U for simultaneously chosen allocation and appreciation is a boundary value, which is the same for the social planner and decentralised case. Variables with * denote the fully optimal solution in what follows.

We next notee the basic property of our approach that a change in climate-friendly preferences, *ceteris paribus,* reduces emissions. If the tax is set too low so that the climate target will not be reached, then a change towards climate-friendly preferences helps to make up for inadequate regulation and closes some of the gap between actual and desired emissions. To see this, let utility be parametrised as in Equation (4), so that clean and dirty consumption are substitutes.

The first-order condition of the consumer is for the parametrised case given by (see Appendix A)

$$\frac{C^M}{D^M} = \Big(\frac{(1-\alpha^M)}{\alpha^M}\Big)^{\frac{\gamma}{(\gamma-1)}} \Big(\frac{1}{1+\tau}\Big)^{\frac{1}{(\gamma-1)}}. \tag{8}$$

Derive this expression with respect to α^M, to find that

$$\frac{C^M/D^M}{\partial \alpha^M} > 0 \qquad \text{if } \gamma > 0. \tag{9}$$

So a marginal positive change in appreciation will reduce emissions, *ceteris paribus.*

Economists often correctly stress the importance of pricing, given inadequate voluntary action. Environmentalists often correctly stress the importance of voluntary action, given inadequate pricing. Standard models with fixed preferences are unable to make sense of voluntary reduction as a change in preferences. The noted basic property of our model reconciles these views, identifying merit of both approaches. It is similar to Perino (2015), who studies the impact of climate campaigns (understood as modifications of the utility function) on aggregate emissions in general equilibrium. Importantly, Perino (2015) shows that the result hinges on total or partial regulation of an economy's emission by a tax or a permit scheme and the emission-intensity of the sectors regulated. This is not our focus, instead we elaborate on the consequences for regulation when the tax itself affects preferences.

In the decentralised case, it cannot be assumed that $\alpha^M = \alpha^{\text{SO}}$, that is that the relative appreciation of clean and dirty goods equals the socially optimal appreciation (with the externality fully corrected). Instead, suppose for idealised first-best policy, the government had a policy instrument to adjust appreciation ϵ : information campaigns or education (abstracting from the costs incurred for the government to do so, just as typically one abstracts from the transaction costs of levying taxes etc.). This would act on appreciation as follows:

$$U((\alpha + \epsilon)C, (1 - (\alpha + \epsilon))D). \tag{10}$$

In this case, trivially, the first-best appreciation-adjusting policy is $\epsilon = \alpha^{\text{SO}} - \alpha^M$. In public finance, it is standard to rule out inidividualised lump-sum transfers to characterise feasible tax policy. In the remainder of this section, we think of such an unrealistic instrument as the analogue of individualised lump-sum taxes and hence rule it out.

3.3 Pricing that changes preferences

Thus far, we have examined how prices and preferences can separately contribute to achieving a specific environmental target. However, our analysis has assumed separability between prices and preferences – prices did not affect preferences. Bowles and Polania-Reyes (2012) provides ample evidence against this assumption, so we move to consider the case when, as in Lanz et al. (2018) (see Section 2.1), environmental prices can change preferences by crowding-in or -out. We model this examining the properties of *any* instrument that influences both the relative price and appreciation in a static setting under certainty. We model the shift in appreciation caused by a carbon price for *target-compatible* regulation. Let $f(\alpha^M, \tau)$, defined over $[0, 1] \times \mathbb{R}^+$ be a smooth function that describes how taxes act on preferences of specific goods. The consumer problem is:

$$\max_{C,D} U(f(\alpha^M, \tau)C, (1 - f(\alpha^M, \tau)D) \tag{11}$$

subject to the budget constraint:

$$p_C C + (p_D + \tau)D = w + L. \tag{12}$$

We assume throughout that $\alpha^M < \alpha^{SP}$, so that if $f(\alpha^M, \tau) > \alpha^M$, appreciation is crowded-in by environmental pricing and crowded-out otherwise. Further, we assume that $0 < f(\alpha^M, \tau) < 1$ to ensure meaningful solutions.

Let U_X, U_Y denote the derivative with respect to the first and second component of the utility function, respectively. The first-order condition for the consumer is:

$$\big(1 - f(\alpha^M, \tau)\big)\, U_Y = \frac{1}{p_C}(p_D + \tau) f(\alpha^M, \tau) U_X \tag{13}$$

We assume, with a broad macroeconomic price signal in mind, that the effects on preferences are small. So, as an approximation, we give f an explicit linear form $f(\alpha^M, \tau) = \alpha^M + \beta\tau$, to allow representation of the crowding effects as a single parameter β, as does Bowles (2008). That is, for classical separability of prices and preferences, $\beta = 0$. Call β the 'crowding-in constant'. In this case, equation (13) can be rearranged to the following implicit expression for a second-best tax:

$$\frac{1}{p_C}(p_D + \tau^{SB}) = \frac{1 - (\alpha^M + \beta\tau^{SB})}{(\alpha^M + \beta\tau^{SB})} z^{SB}(\beta, \tau^{SB}). \tag{14}$$

z^{SB} is the inverse of the marginal rate of utility substitution, however it may in general depend on β and τ. For the parametrisation of utility chosen (see Equation (4), and an environmental target reflected in $\widetilde{C}, \widetilde{D}$, it is given by:

$$z^{SB} = U_Y)^{SB}/(U_X)^{SB} = \big(\frac{1 - \alpha^M + \beta\tau}{\alpha^M + \beta\tau}\big)^{\gamma-1}\big(\frac{\widetilde{D}}{\widetilde{C}}\big)^{\gamma-1}$$

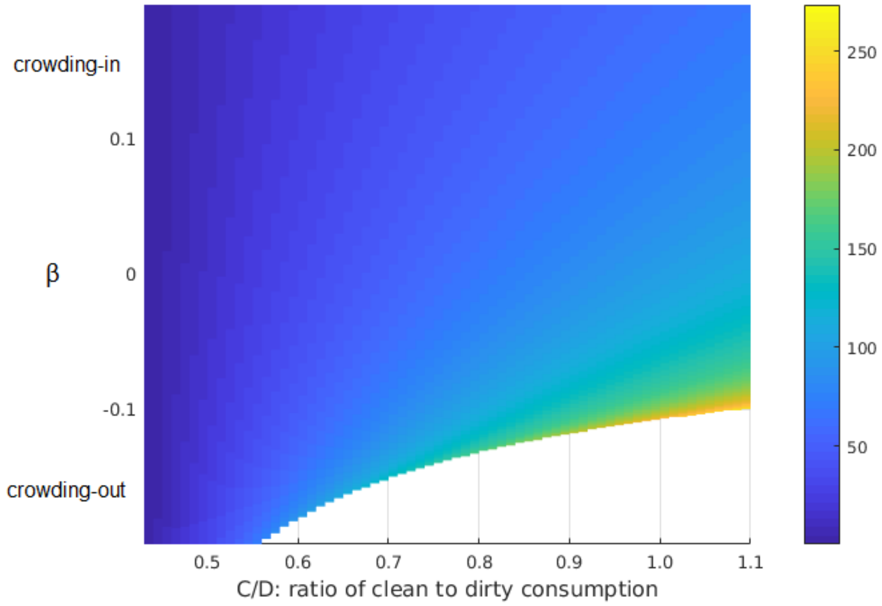
Therefore, the second-best tax is given implicitly by:

$$\tau^{SB} = \frac{1}{p_C}\big(\frac{1 - (\alpha^M + \beta\tau^{SB})}{(\alpha^M + \beta\tau^{SB})}\big)^{\gamma}\big(\frac{\widetilde{D}}{\widetilde{C}}\big)^{\gamma-1} - p_D. \tag{15}$$

Similarly, if α^M denotes fixed appreciation of consumers, and z^{FA} the inverse of the marginal rate of utility substitution, it can be shown that (Appendix A)

$$\tau^{FA} = p_C \frac{1 - \alpha^M}{\alpha^M} z^{FA} - p_D. \tag{16}$$

Figure 2: Target-achieving CO_2 price as a function of the strength of crowding-in social preferences β and of the environmental target



Parametrisation with $p_D = p_C = 1, \gamma = 1/2, \alpha = 0.3$, additionally varying the environmental target. Prices need to be higher with crowding-in and more ambitious environmental targets. Solutions in the right bottom corner do not exist. Prices normalised to 100 for $\beta = 0, C/D = 1$.

with $z^{FA} = \left(\frac{1-\alpha^M}{\alpha^M}\right)^{\gamma-1}\left(\frac{\widetilde{D}}{\widetilde{C}}\right)^{\gamma-1}$ for the standard parametrisation.

Note that a solution to Equation (15) need not have a solution in general for any environmental target if $\beta < 0$. The reason is that, in the case of crowding-out, a change in relative price in favour of the clean good may be counteracted by a larger crowding-out effect. [13] We simulated the solution to Equation (15) numerically for various values of γ; see Figure 2 for an illustration for $\gamma = 0.5$). Our simulation indicates that, even before reaching the limits of an economically meaningful outcome, there are no solutions to Equation (15) for negative β and ambitious environmental targets. See also (Konc et al., 2021) for a calibrated simulation of the impact of a tax on a socially undesirable good on interdependent preferences in a context of social networks.

Mattauch and Hepburn (2016) and Lanz et al. (2018) state that the carbon tax needs to be adjusted in the presence of crowding-out: Suppose preferences are endogenous to a carbon price. If there is crowding-out, then

[13] Further, there are limits to economically meaningful solutions given by $\alpha + \beta\tau = 1$ and $\alpha + \beta\tau = 0$, as the consumer only derives utility from the clean or dirty good respectively.

the carbon price to achieve a target needs to be higher than if they were exogenous. This can now be proved.

Proposition 2. *Let utility be parametrised as above. When $\beta < 0$, ("crowding-out") the second-best carbon price τ^{SB} needs to be higher than the conventional first-best price τ^{FA} (fixed appreciation) to achieve the desired level of mitigation and lower if $\beta > 0$.*

Proof. Compare the parametrised version of Equations (14) and (16) for a positive carbon price. For $\beta > 0$,

$$\frac{1-\alpha^M}{\alpha^M} > \frac{1-(\alpha^M+\beta\tau^{SB})}{(\alpha^M+\beta\tau^{SB})} \tag{17}$$

as long as the second-best tax is positive. Since the same environmental target is to be achieved and noting $\gamma > 0$, this implies $\tau^{FA} > \tau^{SB}$. The inequality is reversed if $\beta < 0$.

□

Further, since we have defined a social optimum in Subsection 3.1, we can compare the target-compatible carbon prices with optimal appreciation with the case in which appreciation is endogenous. The next result characterises how the optimal case relates to Proposition 2.

Proposition 3. *Let utility by parametrised as above. When $\beta > 0$ ("crowding in"), there exists a unique value $\beta^\dagger$ such that the second-best tax equals the socially optimal carbon price τ^{FB} (with optimal appreciation). The second-best tax is higher than the socially optimal price if $\beta < \beta^\dagger$ and* vice versa.

Given that there is crowding-in, there will be just one value of how much is crowded in which makes it possible for the tax to both achieve the target and at the same time crowd-in enough to achieve optimal appreciation.

Proof. See Appendix. □

Bowles (2016) argues that the *design of a monetary incentive itself* can lead to changes in the degree of crowding-in or -out of social preferences that comes with the incentive. In the above notation, this would imply that β is a function of the level and design of a carbon price. The design features that may influence β are, for instance, likely to be related to political messaging around the policy reform: good communication of pollution regulation will make citizens feel empowered, not patronised (as exemplified for example by the success of the Irish plastic bag tax (Convery et al., 2007; Bowles, 2016)). Further, high trust in politicians correlates with higher carbon prices (Klenert et al., 2018; Rafaty, 2018). So it could be that it is "bad news"

(Bowles, 2016) when trust is low and a government announces plans for an environmental tax, while it is not when citizens trust their politicians.

For the case of optimal regulation – that is achieving the optimal amount of climate damage – rather than target-compatible, regulation, we obtain the following:

Proposition 4. *When taxes set to maximise utility do not crowd in low-carbon preferences enough (too much), the government should adjust tax levels upwards (downwards) to achieve the social optimum.*

Proof. See Appendix. □

By symmetry, when taxes crowd out low-carbon preferences, the optimum can only be reached by a tax when there was too much appreciation of the clean good prior to taxation – though admittedly this is of little policy relevance. Further work could explore the deviation of the conventional optimal tax when it affects appreciation. It could also explore the properties of the second-best tax, that is whether it is set above or below the Pigou level.[14]

Furthermore, in this setting one can consider an instrument designed to shift appreciation, such as awareness campaigns or educatory measures, but which will come at some opportunity cost of consumption. However, in the special case of crowding-in, in which a price that changes preferences can restore the full social optimum (Proposition 3), such an instrument would be an inferior alternative to the price signal due to its opportunity cost and there is no additional use of it. When the price signal is suboptimally set or cannot reach the full optimum, there are potential welfare gains from an awareness campaign, but only under the trade-off between beneficial changes in appreciation and income reduction (see Appendix A.5 for a formal sketch).

4 Application to transport and food choices and normative adequacy

We illustrate how the idea that regulation can change preferences leads to new policy conclusions by two further applications: the role of urban transport infrastructure, and health benefits from choosing low-carbon consumption options. We then explain the broad normative basis for our approach.

4.1 Urban transport infrastructure

As noted in Subsection 2.1, Weinberger and Goetzke (2010) showed that the built urban environment can determine propensity for car ownership in the

[14]See Pigou (1920). In this context, a Pigouvian price is the price at which the external cost is fully internalised with the impact of prices on preferences being accounted for.

long run. If this holds, an evaluation of transport infrastructure that ignores the impact on preferences, focussing instead on price impacts, will lead to inefficient policy, as it will understate the benefits of shifting preferences that facilitate low-carbon transport.

Assume that the effect of infrastructure on preferences only occurs in the future, because preferences about mobility options are formed in the long term. Consider a two period model. For simplicity, decision-makers optimise their mobility behaviour for the two periods separately. We think of this as two distinct generations, those taking urban transport decisions now and those who will live in future cities. We study a policy instrument that shifts prices and preferences in the second period, but needs financing in the first period.

Period 1:

$$\max_{C_1,D_1} U(\alpha_1^M C_1, (1-\alpha_1^M)D_1) \tag{18}$$

subject to

$$p_{c1}C_1 + p_{d1}D_1 = w_1 - T \tag{19}$$

Period 2:

$$\max_{C_2,D_2} U((\alpha_2^M + g(T))C_2, (1-(\alpha_2^M + g(T)))D_2) \tag{20}$$

subject to

$$p_{c2}(T)C_2 + p_{d2}D_2 = w_2 \tag{21}$$

Here we assume that the consumers ignore a consumption externality about urban environmental quality E.

Infrastructure investment needs to be financed in the first period, but will change both relative prices and appreciation in the second period. The latter effect is represented as a function $g(T) > 0$ with $g' > 0$. There is a trade-off between consumption losses due to infrastructure financing and correcting both externality and appreciation. We now compare two different models by the following statement:

Proposition 5. *Assume clean and dirty mobility options are ordinary goods. For a given level of low-carbon infrastructure financing T, if $g > 0$, that is if infrastructure locks-in mobility preferences, in the second period, the social marginal value of clean (dirty) consumption is higher (lower) and hence more infrastructure investment is warranted.*

Proof. See Appendix. □

Appropriate transport infrastructure can be assumed to raise the share of low-carbon transport due to lower relative prices and to lock-in of preferences for low-carbon transport, as exhibited by Weinberger and Goetzke (2010).

We conclude that if there is suboptimal appreciation for the second-period, then infrastructure investment is more important than typically assumed.[15]

Section 2.1 also noted that public health policies can shape preferences to help people to make healthier and more environmentally-beneficial choices (Hawkes et al., 2015). The examples of Woodcock et al. (2009) and Springmann et al. (2016) highlight that significant welfare gains could be achieved by increasing active travel and reducing the fraction of animal-sourced foods, since such changes reduce both emissions and obesity-related diseases. Habitual car-driving for short trips or consuming large quantities of red meat beyond dietary requirements is structurally similar to smoking or drink-driving, generating both an "internality" (harm to the individual created by costs on future health) and an "externality" (harm to others). Such behaviour, taken together with stated preferences about the importance of health, indicates that citizens entertain different, conflicting, preferences about health outcomes, and for the purpose of decarbonisation only a subset of these preferences are helpful.

Two approaches could be useful to elucidate adequate policy for these examples. First, in behavioural economics, conflicting long- and short-term preferences are standardly modelled with time-inconsistent preferences (Laibson, 1997). Quasi-hyperbolic discounting can be used to make sense of the idea that while decision-makers have a long-term preference for staying in good health, they have a short-term preference for unhealthy food or inactive travel behaviour. One could hence combine the model of a quasi-hyperbolic decision maker with an environmental externality to study first and second-best policy.

Alternatively, one could extend the model of Section 3 and account for the possibility of preference changes over mobility and health choices in order to model the reduction of diseases through mitigation policy. One reason to prefer this approach over the first one suggested here is that heterogeneity in preferences with regard to red meat consumption and car-driving (Gao et al., 2017; McLaren, 2007; Ogden et al., 2014; Woodcock et al., 2009) are difficult to represent credibly *only* through differences in time preference rates.

Formally, consider a decision-maker (in a static context, for simplicity) whose utility also depends on his health:

$$U(\alpha C, (1-\alpha)D, H, E) \tag{22}$$

subject to $C + D = w$. Here, let C, D denote clean and dirty consumption respectively, α the appreciation of the respective option, H health and E environmental quality, with consumers to some degree ignoring effects of

[15] We abstracted from a pricing instrument such as a city toll here; however, the most relevant situation for changing appreciation by infrastructure may be when such a price signal to improve environmental outcomes is missing (see Siegmeier (2016)).

their choices on environmental quality and health. Assume the following relationships between the consumption options, health and the environment:

$$E = f(D) \quad \text{and} \quad H = g(C, D) \tag{23}$$

with $f'(\cdot) < 0$, meaning that dirty consumption influences environmental quality negatively. For the case of transport, the shape of g is usefully approximated by $g(C, D) = g(C)$ with $g'(C) > 0, g''(C) < 0$ because there is some evidence that additional active travel does not crowd-out other types of physical activity (Laeremans et al., 2017). The shape of g will be more complicated, however, for the case of diets. Given that current outcomes are suboptimal, since too much greenhouse gases are emitted and health benefits not taken sufficiently into account – a partial "internality"– , one could calculate the societal benefit of a change in preferences, represented by α, similar to the model of Section 3. For example, if a tax on meat consumption, say, crowded in- or out intrinsic motivation to eat a plant-based diet, the above formal analysis would apply, but with an additional change to utility gained from health.

4.2 Normative and policy implications

If society does not debate how preferences are formed, and subjects preference formation to explicit democratic control, preferences are at risk of developing without clarity about what is at stake, and with a risk that they are shaped to profit specific special interest groups rather than society as a whole.[16] Failing to discuss possible shifts in preferences also arguably places greater weight upon the status quo. Given the importance of changes in preferences for major social transitions, such as that necessary to a net zero carbon economy, it is important to account for how value changes interact with policy instruments.

Endogenising preferences makes welfare analysis more challenging, but progress can be made in one of three ways (see Mattauch and Hepburn, 2016, for a treatment of the relative merits of these approaches). We noted that our above analysis relies on the ability to compare utilities for different preferences in a meaningful way. One way to do this is to observe the existence of 'meta-preferences' beyond the first-order level of preferences which are endogenous (see e. g. Sen, 1977). A consumer may, for instance, like herself more when intrinsically motivated to protect the environment. Another approach is to proceed by assuming that utility is cardinal, rather than ordinal. When the intensity of utility changes is taken into account,

[16] See also Bowles (2016), Fehr and Hoff (2011), Hoff and Stiglitz (2016) and Thaler and Sunstein (2008) for related views. Fehr and Hoff (2011) also refute the claim that endogenising preferences introduces too many degrees of freedom. Further, see Epstein and Robertson (2015) for an example of the potential of search engines to influence preferences.

this gives a unit by which one can compare different preferences and their corresponding utility functions (see e. g. Fleurbaey, 2009).

A third possibility, which we did not apply above, is to conduct welfare analysis of endogenous preferences with the equivalent income approach, but with given reference prices (see Fleurbaey and Tadenuma, 2014; Fleurbaey, 2016, for more extensive discussions). This approach relies on money-metric utilities and is related to compensating and equivalent variation, the classic method of doing welfare analysis with revealed preferences as the only source of information. Converting different preferences (also those changed by policy) into different income levels, via money-metric utilities with reference prices, is another way to compare them.[17]

In this article, we have not needed to specify whether our utility functions have an ordinal or cardinal interpretation, as long as a comparison in terms of appreciation is possible. Our approach, however, relies on the assumption that a single parameter can be used to translate relevant aspects of preferences into "appreciation" for low-carbon consumption – and, crucially, we assume that the relationship between goods and their appreciation is cardinal and that appreciation can be given a numerical value. This assumption, while heroic, seems neutral with respect to the different normative views of individual well-being.

Finally, two implications of our study could be examined in future work. One is the question of deliberate sequencing of policy (Meckling et al., 2017; Pahle et al., 2018): Should a price signal or an awareness campaign come first to foster further decarbonisation? Or should they be introduced at the same time? Scattered evidence in environmental psychology indicates that the sequence of behavioural interventions matters for the success of behaviour change with respect to mobility decisions (Gatersleben and Appleton, 2007; Bamberg, 2013). So if value change facilitates the introduction of relative price changes and brings down the required level of carbon pricing, it is possible that the timing and coordination between interventions on prices and preferences also matters for the efficiency of environmental policy.

Four related questions about the political economy of value changes arises. First, for any actor in government, is it easier to enact policy that changes relative prices by taxes, subsidies or bans, or is it easier to enact policy that changes relative preferences by information, persuasion and education? Some governments run awareness campaigns about the environment, although governments are often incapable of setting carbon prices anywhere

[17]von Weizsäcker (1971, 2005, 2013) has pursued a different line of thinking about endogenous preferences, developing criteria under which preferences can change and Paretian welfare economics is still feasible. He considers preferences as "adaptive", defining them as follows. "[I]ndividuals have a tendency to value their present position or situation higher relative to alternatives than they would, if their present position or situation were a different one. We also may call this preference conservatism: a tendency of agents to stick to the place where they are." (von Weizsäcker, 2013, p. 14)

near target-adequate levels (Stiglitz and Stern, 2017), for example because public support of carbon prices is low (Klenert et al., 2018). This suggests that it could in some situations be 'politically' easier, at least at the margin and relative to the status quo, to change preferences than relative prices.

Second, bringing in carbon pricing and returning revenue to consumers could be a way of beginning to change preferences, but then, once they have begun to change, the uses of revenues might become more open. Besley and Persson (2019) already study how the share of "environmentalists" co-evolves with taxes on emissions in a rational electoral competition framework. A political economy approach to our context could consider a government that faces political obstacles to price-changing and appreciation-changing policies.

Third, our approach underlines that preferences and the economic context in which they are displayed cannot be separated (Bowles and Polania-Reyes, 2012; Bowles, 2016). Yet, for specific applications, such as taxes on fuel or meat, our article does not address the question whether changes in preferences could be irreversible. That is, after a societal transition, regulation may be needed to a lesser extent. For example, in a society of vegetarians bound together by strong social norms against eating meat, carbon prices on animal products are unlikely to be a political priority.

Finally, while the present article solely focussed on changes in consumer behaviour, changes in business norms in response to environmental regulation and targets also seem underexplored in environmental public economics. For example, many consumer good companies have announced to be carbon neutral by a certain date in response to global climate targets. Some companies are under pressure from divestment campaigns and other firms apply internal carbon prices higher than current national carbon prices to steer their operations. Further, for financial institutions, increased disclosure of climate alignment required by governments and a published comparison between them may result in norm changes on the importance of how much they contribute to a low-carbon transition. Such repercussions on the "preferences of businesses" of climate policy-making seems unexamined from the point of optimal policy evaluation.

5 Conclusion

Policy-induced changes in consumers' preferences are relevant for decarbonisation, given significant empirical evidence that preferences do change with policies – intentionally or unintentionally. Our understanding of climate change mitigation policy would be enhanced if relevant effects were taken into account in economic models. We establish the following results about instrument design under the assumption that preferences are endogenous policy: First, if a climate target is to be achieved and carbon pricing is insufficient, a change in consumers' preferences towards low-carbon preferences helps to achieve the target. Even under the assumption of separability between prices and preferences, strategies to change prices and preferences may be complementary. Second, when the introduction of a carbon tax changes consumers' preferences, not merely relative prices, the target-compatible carbon price must be adjusted by the size of this effect. Third, when low-carbon infrastructure leads to the formation of low-carbon preferences over time, we should account for the additional climate-protection value of investing in such infrastructure. Finally, the potential for health gains through reductions in obesity-related diseases from low-carbon diets and active urban travel – also an area where policy shapes preferences – provide an additional reason for supporting the evolution of preferences towards healthy eating and mode choice. To deliver on the temperature targets of the Paris agreement on climate change, public policy should consider whether these effects yield additional levers for accelerating the societal transition to the carbon-neutral economy.

References

Abel, A. B. (1990). Asset prices under habit formation and catching up with the joneses. *American Economic Review*, 80(2):38–42.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Akerlof, R. (2017). Value formation: The role of esteem. *Games and Economic Behavior*, 102:1–19.

Alesina, A. and Fuchs-Schündeln, N. (2007). Good-bye Lenin (or not?): The effect of communism on people's preferences. *The American Economic Review*, 97(4):1507–1528.

Algan, Y. and Cahuc, P. (2010). Inherited trust and growth. *The American Economic Review*, 100(5):2060–2092.

Allcott, H., Mullainathan, S., and Taubinsky, D. (2014). Energy policy with externalities and internalities. *Journal of Public Economics*, 112:72–88.

Allen, M. W., Wilson, M., Ng, S. H., and Dunne, M. (2000). Values and beliefs of vegetarians and omnivores. *The Journal of Social Psychology*, 140(4):405–422.

Andersson, J. J. (2019). Carbon taxes and co_2 emissions: Sweden as a case study. *American Economic Journal: Economic Policy*, 11(4):1–30.

Bamberg, S. (2013). Applying the stage model of self-regulated behavioral change in a car use reduction intervention. *Journal of Environmental Psychology*, 33:68–75.

Becker, G. S. and Mulligan, C. B. (1997). The endogenous determination of time preference. *The Quarterly Journal of Economics*, 112(3):729–758.

Bénabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Bernheim, B. D. and Rangel, A. (2009). Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, 124(1):51–104.

Besley, T. and Persson, T. (2019). Jeea-fbbva lecture 2017: The dynamics of environmental politics and values. *Journal of the European Economic Association*, 17(4):993–1024.

Bezin, E. (2015). A cultural model of private provision and the environment. *Journal of Environmental Economics and Management*, 71:109–124.

Bezin, E. (2019). The economics of green consumption, cultural transmission and sustainable technological change. *Journal of Economic Theory*, 181:497–546.

Bezin, E. and Ponthière, G. (2019). The tragedy of the commons and socialization: Theory and policy. *Journal of Environmental Economics and Management*, 98:102260.

Binder, M. (2010). *Elements of an evolutionary theory of welfare: assessing welfare when preferences change.* Routledge, Abingdon.

Birch, L. L. (1999). Development of food preferences. *Annual Review of Nutrition*, 19(1):41–62.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, 36(1):75–111.

Bowles, S. (2008). Policies designed for self-interested citizens may undermine" the moral sentiments": Evidence from economic experiments. *Science*, 320(5883):1605–1609.

Bowles, S. (2016). *The moral economy. Why good incentives are no substitute for good citizens.* Yale University Press, New Haven.

Bowles, S. and Hwang, S.-H. (2008). Social preferences and public economics: Mechanism design when social preferences depend on incentives. *Journal of Public Economics*, 92(8-9):1811–1820.

Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50(2):368–425.

Brennan, T. J. (2006). green preferences as regulatory policy instrument. *Ecological economics*, 56(1):144–154.

Cavatorta, E. and Groom, B. (2020). Does deterrence change preferences? evidence from a natural experiment. *European Economic Review*, 127:1–16.

Convery, F., McDonnell, S., and Ferreira, S. (2007). The most popular tax in europe? lessons from the irish plastic bags levy. *Environmental and resource economics*, 38(1):1–11.

Corner, A., Markowitz, E., and Pidgeon, N. (2014). Public engagement with climate change: the role of human values. *Wiley Interdisciplinary Reviews: Climate Change*, 5(3):411–422.

Costa, F. and Gerard, F. (2018). Hysteresis and the social cost of corrective policies: evidence from a temporary energy saving program. NBER Working Paper No. 24608.

Dasgupta, P., Southerton, D., Ulph, A., and Ulph, D. (2016). Consumer behaviour with environmental and social externalities: implications for analysis and policy. *Environmental and Resource Economics*, 65(1):191–226.

Daube, M. and Ulph, D. (2016). Moral behaviour, altruism and environmental policy. *Environmental and Resource Economics*, 63(2):505–522.

Deckers, T., Falk, A., Kosse, F., and Schildberg-Hörisch, H. (2020). Socio-economic status and inequalities in childrens iq and economic preferences. *Journal of Political Economy*, 128:434–67.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–372.

D'Haultfœuille, X., Durrmeyer, I., and Février, P. (2016). Disentangling sources of vehicle emissions reduction in france: 2003–2008. *International Journal of Industrial Organization*, 47:186–229.

Epstein, R. and Robertson, R. E. (2015). The search engine manipulation effect (seme) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 112(33):E4512–E4521.

Farhi, E. and Gabaix, X. (2020). Optimal taxation with behavioral agents. *American Economic Review*, 110(1):298–336.

Fehr, E. and Hoff, K. (2011). Introduction: Tastes, castes and culture: the influence of society on preferences. *The Economic Journal*, 121(556):F396–F412.

Fleurbaey, M. (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, 47(4):1029–1075.

Fleurbaey, M. (2016). Equivalent income. In *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press.

Fleurbaey, M. and Tadenuma, K. (2014). Universal social orderings: An integrated theory of policy evaluation, inter-society comparisons, and interpersonal comparisons. *Review of Economic Studies*, 81(3):1071–1101.

Frey, B. S. and Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review*, 87(4):746–755.

Frey, B. S. and Stutzer, A. (2008). Environmental morale and motivation. In Lewis, A., editor, *The Cambridge Handbook of Psychology and Economic Behaviour*, pages 406–28. Cambridge University Press, Cambridge.

Fuchs-Schündeln, N. and Schündeln, M. (2015). On the endogeneity of political preferences: Evidence from individual experience with democracy. *Science*, 347(6226):1145–1148.

Gao, J., Helbich, M., Dijst, M., and Kamphuis, C. B. (2017). Socioeconomic and demographic differences in walking and cycling in the netherlands: How do these translate into differences in health benefits? *Journal of Transport & Health*, 6:358–365.

Gatersleben, B. and Appleton, K. M. (2007). Contemplating cycling to work: Attitudes and perceptions in different stages of change. *Transportation Research Part A: Policy and Practice*, 41(4):302–312.

Gintis, H. (1974). Welfare criteria with endogenous preferences: the economics of education. *International Economic Review*, 15(2):415–430.

Götschi, T., Tainio, M., Maizlish, N., Schwanen, T., Goodman, A., and Woodcock, J. (2015). Contrasts in active transport behaviour across four countries: How do they translate into public health benefits? *Preventive Medicine*, 74:42–48.

Grinblatt, M., Keloharju, M., and Ikäheimo, S. (2008). Social influence and consumption: Evidence from the automobile purchases of neighbors. *The review of Economics and Statistics*, 90(4):735–753.

Hahn, R., Metcalfe, R., et al. (2016). The impact of behavioral science experiments on energy policy. *Economics of Energy & Environmental Policy*, 5(2):27–44.

Hawkes, C., Smith, T. G., Jewell, J., Wardle, J., Hammond, R. A., Friel, S., Thow, A. M., and Kain, J. (2015). Smart food policies for obesity prevention. *The Lancet*, 385(9985):2410–2421.

Hoff, K. and Stiglitz, J. E. (2016). Striving for balance in economics: Towards a theory of the social determination of behavior. *Journal of Economic Behavior & Organization*, 126:25–57.

Hopkins, D. (2016). Can environmental awareness explain declining preference for car-based mobility amongst generation y? a qualitative examination of learn to drive behaviours. *Transportation Research Part A: Policy and Practice*, 94:149–163.

Hunecke, M., Blöbaum, A., Matthies, E., and Höger, R. (2001). Responsibility and environment: Ecological norm orientation and external factors in the domain of travel mode choice behavior. *Environment and Behavior*, 33(6):830–852.

IPCC (2013). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change. Stocker, Thomas et al. (ed).* Cambridge University Press.

IPCC (2018). *Global warming of* 1.5 °*C. An IPCC Special Report on the impacts of global warming of 1.5C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V. and Zhai, P. and Prtner, H.O. and Roberts, D. and Skea, J. and Shukla, P.R. and Pirani, A. and Moufouma-Okia, W. and P'ean, C. and Pidcock, R. and Connors, S. and Matthews, J.B.R. and Chen, Y. and Zhou, X. and Gomis, M.I. and Lonnoy, E. and Maycock, T. and Tignor, M. and Waterfield, T.].* World Meteorological Organization.

Jacobsen, G. D., Kotchen, M. J., and Vandenbergh, M. P. (2012). The behavioral response to voluntary provision of an environmental public good: Evidence from residential electricity demand. *European Economic Review*, 56(5):946–960.

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin, London.

Klenert, D., Mattauch, L., Combet, E., Edenhofer, O., Hepburn, C., Rafaty, R., and Stern, N. (2018). Making carbon pricing work for citizens. *Nature Climate Change*, 8:669–677.

Koessler, A.-K. and Engel, S. (2019). Policies as information carriers: How environmental policies may change beliefs and consequent behavior.

Konc, T., Savin, I., and van den Bergh, J. C. (2021). The social multiplier of carbon taxation. *Journal of Environmental Economics and Management*, 105:102396.

Laeremans, M., Gotschi, T., Dons, E., Kahlmeier, S., Brand, C., de Nazelle, A., Gerike, R., Nieuwenhuijsen, M., Raser, E., Stigell, E., et al. (2017). Does an increase in walking and cycling translate into a higher overall physical activity level? *Journal of Transport & Health*, 5:S20.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2):443–477.

Lanz, B., Wurlod, J.-D., Panzone, L., and Swanson, T. (2018). The behavioral effect of pigovian regulation: Evidence from a field experiment. *Journal of Environmental Economics and Management*, 87:190–205.

Levy, D. T., Boyle, R. G., and Abrams, D. B. (2012). The role of public policies in reducing smoking: the minnesota simsmoke tobacco policy model. *American Journal of Preventive Medicine*, 43(5):S179–S186.

Lichtenstein, S. and Slovic, P. (2006). *The construction of preference*. Cambridge University Press.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65(3):272–292.

Loewenstein, G. and Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics*, 107(2):573–597.

Malmendier, U. and Nagel, S. (2011). Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, 126(1):373–416.

Mattauch, L. and Hepburn, C. (2016). Climate policy when preferences are endogenousand sometimes they are. *Midwest Studies In Philosophy*, 40(1):76–95.

Mattauch, L., Ridgway, M., and Creutzig, F. (2016). Happy or liberal? Making sense of behavior in transport policy design. *Transportation Research Part D: Transport and Environment*, 45:64–83.

McLaren, L. (2007). Socioeconomic status and obesity. *Epidemiologic Reviews*, 29(1):29–48.

Meckling, J., Sterner, T., and Wagner, G. (2017). Policy sequencing toward decarbonization. *Nature Energy*, 2(12):918.

Mons, U., Nagelhout, G. E., Allwright, S., Guignard, R., van den Putte, B., Willemsen, M. C., Fong, G. T., Brenner, H., Pötschke-Langer, M., and Breitling, L. P. (2013). Impact of national smoke-free legislation on home smoking bans: findings from the international tobacco control policy evaluation project europe surveys. *Tobacco Control*, 22:e2–e9.

Mueller, N., Rojas-Rueda, D., Cole-Hunter, T., de Nazelle, A., Dons, E., Gerike, R., Götschi, T., Panis, L. I., Kahlmeier, S., and Nieuwenhuijsen, M. (2015). Health impact assessment of active transportation: a systematic review. *Preventive Medicine*, 76:103–114.

Mueller, N., Rojas-Rueda, D., Salmon, M., Martinez, D., Brand, C., de Nazelle, A., Gerike, R., Gotschi, T., Iacorossi, F., Panis, L. I., et al. (2017). Health impact assessment of cycling network expansions in european cities. *Journal of Transport & Health*, 5:S9–S10.

Newman, P. and Kenworthy, J. (1999). *Sustainability and cities: overcoming automobile dependence*. Island Press, Washington, DC.

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W. N., Arrow, K. J., Barrett, S., Carpenter, S., et al. (2016). Social norms as solutions. *Science*, 354(6308):42–43.

Ogden, C. L., Carroll, M. D., Kit, B. K., and Flegal, K. M. (2014). Prevalence of childhood and adult obesity in the united states, 2011-2012. *Jama*, 311(8):806–814.

O'Hara, S. U. and Stagl, S. (2002). Endogenous preferences and sustainable development. *The Journal of Socio-Economics*, 31(5):511–527.

Pahle, M., Burtraw, D., Flachsland, C., Kelsey, N., Biber, E., Meckling, J., Edenhofer, O., and Zysman, J. (2018). Sequencing to ratchet up climate policy stringency. *Nature Climate Change*, 8(10):861–867.

Perino, G. (2015). Climate campaigns, cap and trade, and carbon leakage: Why trying to reduce your carbon footprint can harm the climate. *Journal of the Association of Environmental and Resource Economists*, 2(3):469–495.

Perino, G., Panzone, L. A., and Swanson, T. (2014). Motivation crowding in real consumption decisions: Who is messing with my groceries? *Economic Inquiry*, 52(2):592–607.

Pigou, A. C. (1920). *The economics of welfare*. Palgrave Macmillan, Basingstoke.

Pollak, R. A. (1970). Habit formation and dynamic demand functions. *Journal of political Economy*, 78(4, Part 1):745–763.

Pollak, R. A. (1978). Endogenous tastes in demand and welfare analysis. *The American Economic Review*, 68(2):374–379.

Rabl, A. and De Nazelle, A. (2012). Benefits of shift from car to active transport. *Transport policy*, 19(1):121–131.

Rafaty, R. (2018). Perceptions of corruption, political distrust, and the weakening of climate policies. *Global Environmental Politics*, 18:106–129.

Rivers, N. and Schaufele, B. (2015). Salience of carbon taxes in the gasoline market. *Journal of Environmental Economics and Management*, 74:23–36.

Roos, M. (2018). Values, attitudes and economic behavior. Ruhr Economic Papers 777.

Rozin, P. and Vollmecke, T. A. (1986). Food likes and dislikes. *Annual review of nutrition*, 6(1):433–456.

Russell, W. S. and Zepeda, L. (2008). The adaptive consumer: shifting attitudes, behavior change and csa membership renewal. *Renewable Agriculture and Food Systems*, 23(2):136–148.

Ryder, H. E. and Heal, G. M. (1973). Optimal growth with intertemporally dependent preferences. *The Review of Economic Studies*, 40(1):1–31.

Schumacher, I. (2015). The endogenous formation of an environmental culture. *European Economic Review*, 76:200–221.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of social issues*, 50(4):19–45.

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

Seto, K. C., Dhakal, S., Bigio, A., Blanco, H., Delgado, G. C., Dewar, D., Huang, L., Inaba, A., Kansal, A., Lwasa, S., McMahon, J. E., Müller, D. B., Murakami, J., Nagendra, H., and Ramaswami, A. (2014). Human settlements, infrastructure and spatial planning. In Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Seyboth, K., Matschoss, P., Kadner, S., Zwickel, T., Eickemeier, P., Hansen, G., Schloemer, S., and von Stechow, C., editors, *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report* *of the Intergovernmental Panel on Climate Change*, pages 923–1000. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Severen, C. and Van Benthem, A. (2019). Formative experiences and the price of gasoline. NBER Working Paper 26091.

Siegmeier, J. (2016). Keeping pigou on tracks: second-best combinations of carbon pricing and infrastructure provision. MCC Working Paper 2/2016.

Springmann, M., Godfray, H. C. J., Rayner, M., and Scarborough, P. (2016). Analysis and valuation of the health and climate change cobenefits of dietary change. *Proceedings of the National Academy of Sciences*, 113(15):4146–4151.

Steg, L. (2005). Car use: lust and must. instrumental, symbolic and affective motives for car use. *Transportation Research Part A: Policy and Practice*, 39(2-3):147–162.

Stern, N. (2015). *Why are we waiting?: The logic, urgency, and promise of tackling climate change*. MIT Press.

Stigler, G. J. and Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, 67(2):76–90.

Stiglitz, J. E. and Stern, N. (2017). Report of the High-Level Commission on Carbon Prices (World Bank Carbon Pricing Leadership Coalition). World Bank.

Stuber, J., Galea, S., and Link, B. G. (2008). Smoking and the emergence of a stigmatized social status. *Social science & medicine*, 67(3):420–430.

Sulikova, S., van den Bijgaart, I., Klenert, D., and Mattauch, L. (2020). Optimal fuel taxation with suboptimal health choices. INET Working Paper 2020-22.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, New Haven.

Tompson, S., Lieberman, M. D., and Falk, E. B. (2015). Grounding the neuroscience of behavior change in the sociocultural context. *Current Opinion in Behavioral Sciences*, 5:58–63.

Ulph, A. and Ulph, D. (2018). Environmental policy when consumers value conformity. *Journal of Environmental Economics and Management, online first, doi: 10.1016/j.jeem.2018.09.001*.

van den Bijgaart, I. M. (2018). Too slow a change? Deep habits, consumption shifts and transitory tax policy. CESifo working paper No. 6958.

Villacorta, M., Koestner, R., and Lekes, N. (2003). Further validation of the motivation toward the environment scale. *Environment and Behavior*, 35(4):486–505.

von Weizsäcker, C. C. (1971). Notes on endogenous change of tastes. *Journal of Economic Theory*, 3(4):345–372.

von Weizsäcker, C. C. (2005). The welfare economics of adaptive preferences. MPI Collective Goods Preprint 2005/11.

von Weizsäcker, C. C. (2013). Freedom, wealth and adaptive preferences. Working Paper.

Voors, M. J., Nillesen, E. E., Verwimp, P., Bulte, E. H., Lensink, R., and Van Soest, D. P. (2012). Violent conflict and behavior: a field experiment in burundi. *The American Economic Review*, 102(2):941–964.

Watling, H. and Armstrong, K. (2015). Exploring the influences of country-level factors on mature-aged womens drink driving attitudes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 30:57–65.

Weber, E. U. and Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, 60:53–85.

Weinberger, R. and Goetzke, F. (2010). Unpacking preference: How previous experience affects auto ownership in the United States. *Urban studies*, 47(10):2111–2128.

Woodcock, J., Edwards, P., Tonne, C., Armstrong, B. G., Ashiru, O., Banister, D., Beevers, S., Chalabi, Z., Chowdhury, Z., Cohen, A., Franco, O. H., Haines, A., Hickman, R., Lindsay, G., Mittal, I., Mohan, D., Tiwari, G., Woodward, A., and Roberts, I. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. *The Lancet*, 374(9705):1930–1943.

(Online) Appendix

A Proofs

The first-order condition of the social planer for maximising Equation (3) over (2) with respect to C and D is:

$$\frac{U_C}{U_D} = \frac{\alpha U_X}{(1-\alpha)U_Y} = \frac{p_C}{p_D + \xi} \tag{24}$$

where U_X, U_Y denote the derivative with respect to the first and second component of the utility function, respectively. Taken together with the budget constraint, this gives the optimal allocation (for fixed appreciation), which is unique if preferences are convex. For the parameterised case, the optimal ratio of consumption of clean to dirty goods for given appreciation is:

$$\frac{C}{D} = (\frac{(1-\alpha)}{\alpha})^{\frac{\gamma}{(\gamma-1)}} (\frac{p_C}{p_D + \xi})^{\frac{1}{(\gamma-1)}} \tag{25}$$

Now consider the case of optimally regulating the externality in the decentralised case. For a general utility function, the first-order condition of the consumer (Equations (6) and (7)) is:

$$\frac{U_C}{U_D} = \frac{\alpha U_x}{(1-\alpha)U_y} = \frac{p_C}{p_D + \tau}. \tag{26}$$

In the standard treatment with exogenous appreciation ($\alpha^M = \alpha$), by comparison to Equation (24) it can be seen that the standard Pigouvian tax is $\tau = \xi$.

For the parameterised case, the optimal solution for a given appreciation is:

$$\frac{C^M}{D^M} = (\frac{(1-\alpha^M)}{\alpha^M})^{\frac{\gamma}{(\gamma-1)}} (\frac{1}{1+\tau})^{\frac{1}{(\gamma-1)}} \tag{27}$$

A.1 Proof of Proposition 1

Compare Equation (24) to Equation (26), which both yield an allocation of consumption and differ only by the occurrence of ξ and τ. According to the concavity assumption, the maximum α^* is in each case given by:

$$\frac{\mathrm{d}U(\alpha C^*, (1-\alpha)D^*)}{\mathrm{d}\alpha} = U_X C^* - U_Y D^* = 0 \tag{28}$$

i.e., making use of the envelope theorem. In general, α^* depends on ξ and τ respectively, except in degenerate cases. This proves the proposition.

We illustrate the assumptions of this result by the parameterised version of utility: the proposition is true for $0 < \gamma < 0.5$, since only then does the relevant expression depend on the carbon price. For $\gamma > 0.5$, the optimum is at the boundary at $C = (1/p_c)w, \quad \alpha = 1$. This insight is obtained by differentiating the value function of the social planner problem with respect to α :

$$\frac{\partial U(C,D,\alpha)}{\partial \alpha} = [(\alpha C)^\gamma + ((1-\alpha)D)^\gamma]^{\frac{1}{\gamma}-1}[\alpha^{(\gamma-1)}C^\gamma - (1-\alpha)^{(\gamma-1)}D^\gamma] \tag{29}$$

Setting to zero and inserting the optimal solution $\frac{C^*}{D^*}$ yields:

$$\alpha^* = \frac{\Omega}{1+\Omega} \text{ with } \Omega = \Big(\frac{p_C}{p_D+\xi}\Big)^{\frac{\gamma}{(2\gamma-1)}}. \tag{30}$$

In the decentralised case one finds by analogy:

$$\alpha = \frac{\Omega}{1+\Omega} \text{ with } \Omega = \Big(\frac{p_C}{p_D+\tau}\Big)^{\frac{\gamma}{(2\gamma-1)}}. \tag{31}$$

This only characterises a maximum for $\gamma < 0.5$, however, because otherwise the utility function is not concave in C.

A.2 Derivation of target-compatible regulation with fixed preferences

The "target-optimal" tax can be determined by solving the problem given by Equations (6) and (7) with optimal appreciation to give the first-order condition of the consumer:

$$(1-\alpha^{SO})U_Y = \alpha^{SO}\frac{(p_D+\tau)}{p_C}U_X. \tag{32}$$

Recall that a fixed environmental target $\widetilde{D}$ also fixes $\widetilde{C} = \frac{1}{p_C}(w-(p_D+\xi)\widetilde{D})$ and so determines a socially optimal allocation. This in turn determines a pair of derivatives evaluated at this allocation: $(U_X)^{FB}$, $(U_Y)^{FB}$. From Equation (32) one can deduce the "first-best" tax that achieves the target. Let:

$$z^{FB} = (U_Y)^{FB}/(U_X)^{FB} \tag{33}$$

that is the inverse of the marginal utility rate of substitution at the specified allocation. Then:

$$\tau^{FB} = p_C\frac{1-\alpha^{SO}}{\alpha^{SO}}z^{FB} - p_D. \tag{34}$$

Further, if appreciation is not optimal, but given by some value α^M, the analogue case of *fixed appreciation* is

$$\tau^{FA} = p_C\frac{1-\alpha^{M}}{\alpha^{M}}z^{FA} - p_D. \tag{35}$$

A.3 Proof of Proposition 3

We prove that $\tau^{SB}(\beta)$ is monotone as an implicit function and then apply the Intermediate Value Theorem. Equation (15) defines an implicit function $F(\tau^{SB},\beta)=0$, which we assume to be continuously differentiable. Noting

$$\frac{\mathrm{d}\tau^{SB}}{\mathrm{d}\beta} = -\frac{\partial F/\partial\beta}{\partial F/\partial\tau^{SB}} \tag{36}$$

it can be shown, by computing the derivatives on the right-hand side explicitly, that $\frac{\mathrm{d}\tau^{SB}}{\mathrm{d}\beta} < 0$ for $\beta > 0$. Further, note $\tau^{SB}(0) > \tau^{FB}(0)$, because by assumption $\alpha^M < \alpha^{SP}$. To complete the proof by the Intermediate Value Theorem, it remains to show that it is not the case that $\lim_{\beta\to\infty}\tau^{SB}(\beta) > 0$. (Since the function is monotonically decreasing, a limit exists in $[-\infty, \tau^{SB}(0))$.) Assume for contradiction that the limit is a real positive constant c. Then, from Equation (15) in the limit

$$c + p_D = (-1)^{\gamma}(\frac{\widetilde{D}}{\widetilde{C}})^{\gamma-1}. \tag{37}$$

This is a contradiction because the right-hand side is an imaginary number for $0 < \gamma < 1$ while the left-hand side is real.

A.4 Proof of Proposition 4

For the case of *optimal* regulation, assume again that the effect of the tax on appreciation is linear and consider the problem of the consumer:

$$\max_{C,D} U(\alpha^M C, (1-\alpha^M)D) = [((\alpha^M + \beta\tau)C)^{\gamma} + ((1-(\alpha^M+\beta\tau))D)^{\gamma}]^{\frac{1}{\gamma}} \tag{38}$$

subject to

$$p_C C + (1+\tau)p_D D = w. \tag{39}$$

The solution is given by

$$\frac{C^M}{D^M} = (\frac{1-(\alpha^M+\beta\tau)}{\alpha^M+\beta\tau})^{\frac{\gamma}{\gamma-1}}(\frac{p_C}{p_D+\tau})^{\frac{1}{\gamma-1}}. \tag{40}$$

Compare this to the socially optimal solution:

$$\frac{C^*}{D^*} = (\frac{1-\alpha^*}{\alpha^*})^{\frac{\gamma}{\gamma-1}}(\frac{p_C}{p_D+\tau})^{\frac{1}{\gamma-1}} \tag{41}$$

If $\alpha^* - \alpha^M = \beta\xi$ then the tax conventionally set at $\tau = \xi$ (see above) is Pigouvian, since it optimally corrects for the required change of appreciation. Otherwise, depending on the parameter values, there may be excessive or insufficient shifts in appreciation. Hence the optimal tax should account for $\alpha^* - \alpha^M \neq \beta\xi$. So there is just one specific relationship between how much appreciation needs to change and the strength of the crowding-in needed so that the conventional tax level is the optimal (Pigouvian) one. This yields the result, assuming $\alpha^* > \alpha^M$ and $\beta > 0$.

A.5 Awareness campaign: policies that changes preferences at a cost

Furthermore, assume a real-world instrument designed to shift appreciation, such as awareness campaigns or educatory measures $\epsilon(T)$, will come at some opportunity cost of consumption T. We assume appreciation α^M is suboptimal. So consider the following modified consumer problem:

$$U((\alpha^M + \epsilon(T))C, (1-(\alpha^M+\epsilon(T)))D). \tag{42}$$

subject to the budget constraint

$$p_C C + p_D D = w - T. \tag{43}$$

The consumer maximises utility subject to this policy intervention, in which an appreciation change is financed by a lump-sum tax. Suppose the consumer obtains optimal consumption at $C(T)$ and $D(T)$. Then, to maximise welfare, government will maximise indirect utility V with respect to T :

$$V((\alpha^M + \epsilon(T))C(T), (1 - (\alpha^M + \epsilon(T)))D(T)). \tag{44}$$

As a first-order condition for optimal appreciation-changing policy, one then obtains:

$$V_Y/V_X = -\frac{\frac{\partial C}{\partial T}(\alpha^M + \epsilon(T)) + \frac{\partial \epsilon}{\partial T}C(T)}{\frac{\partial D}{\partial T}(1 - (\alpha^M + \epsilon(T))) - \frac{\partial \epsilon}{\partial T}D(T)}. \tag{45}$$

This characterises a trade-off between adjusting appreciation and the cost of such measures. It is structurally independent of the degree to which the externality is uncorrected and whether a target is to be implemented. Further work could explore the quantitative interaction with a suboptimal carbon price, characterising: (a) the rule for how much adjustment of appreciation is optimal given an environmental target that is implemented, (b) if carbon prices are too low relative to some target, how this rule changes quantitatively, (c) welfare analysis of marginal changes to both instruments when both the carbon price and the awareness campaign are suboptimal.

A.6 Proof of Proposition 5

The government optimises in respect to this choice by considering an indirect utility function with arguments $C_1(T), C_2(T), D_1(T), D_2(T)$, taking into account urban environmental quality in the second period and discounting the future. The government's problem is therefore:

$$\max_T V_1(\alpha_1^M C_1, (1-\alpha_1^M)D_1) + \frac{1}{(1+\rho)} V_2((\alpha_2^M + g(T))C_2, (1-(\alpha_2^M + g(T)))D_2, E(D_2)) \tag{46}$$

The trade-off between consumption losses due to infrastructure financing, and correcting both externality and appreciation is then represented by a first-order condition:

$$\alpha_1^M \frac{\partial V_1}{\partial C_1}\frac{\partial C_1}{\partial T} + (1-\alpha_1^M)\frac{\partial V_1}{\partial D_1}\frac{\partial D_1}{\partial T} + \frac{1}{1+\rho}\Big[\frac{\partial V_2}{\partial C_2}\big(g'(T)C_2 + (\alpha_2^M + g(T))\frac{\mathrm{d}C_2}{\mathrm{d}T}\big) + \frac{\partial V_2}{\partial D_2}\big(-g'(T)D_2 + (1-(\alpha_2^M + g(T)))\frac{\partial D_2}{\partial T}\big) + \frac{\partial V_2}{\partial E}\frac{\partial E}{\partial D_2}\frac{\mathrm{d}D_2}{\mathrm{d}T}\Big] = 0. \tag{47}$$

Note that, for the terms in the second period:

$$\frac{\mathrm{d}C_2}{\mathrm{d}T} = \frac{\partial C_2}{\partial T} + \frac{\partial C_2}{\partial p_{C2}}p'_{C2}(T) \quad \text{and} \quad \frac{\mathrm{d}D_2}{\mathrm{d}T} = \frac{\partial D_2}{\partial T} + \frac{\partial D_2}{\partial p_{C2}}p'_{C2}(T) \tag{48}$$

These terms represent the effect of the relative price change on the value of the policy.

The second part of the left-hand side of Equation (47) characterises the value of an amount of investment needed to enhance environmental quality in the second period at its optimum value. If it was the case that $g = 0$ (no effect on preferences), all terms with g or g' in Equation (47) would disappear. Given that the clean good is ordinary, however, the terms multiplying $\frac{\partial V_2}{\partial C_2}$ are all positive, increasing the value of clean consumption of a fixed investment T, while the opposite is the case for the dirty good.