

Working Paper in Economics No. 757

Optimal Prosocial Nudging

Fredrik Carlsson and Olof Johansson-Stenman

Department of Economics, April 2019

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Optimal Prosocial Nudging

Fredrik Carlsson[#] and Olof Johansson-Stenman[§]

Department of Economics, University of Gothenburg

Abstract

While nudges are still mostly associated with affecting individual choices for their own long-run interest, i.e. dealing with *internalities*, they are increasingly used in order to reduce *externalities*, such as environmental consequences. While we are gaining increasing insights into when and how nudges work, much less attention has been given to the normative aspects of nudging as a policy instrument to deal with externalities. We investigate optimal prosocial nudging under a number of different settings in a world where a conventional Pigovian tax can be used to a varying extent. We find that nudges typically only play a limited role when optimal taxes can be implemented. What we denote encouraging moral nudges, i.e. nudges where people's choices are affected by strengthening consumers' moral norms for doing the right thing, are more likely to play a role even when the tax is optimal compared to purely cognitive nudges. In addition, if a nudge better can target the right consumers, then it might also be optimal to use even when an optimal tax can be implemented. We also present decision rules for the optimal size of a nudge when an optimal tax cannot be implemented.

Key-words: nudge, environmental policy, behavior

JEL codes: D90, H21, H23

[#] Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden; Tel.: +46 31 786 4174; E-mail: fredrik.carlsson@economics.gu.se.

[§] Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden; Tel.: +46 31 786 2538; E-mail: olof.johansson@economics.gu.se.

1. Introduction

The conventional use of nudging (Thaler and Sunstein, 2008) is in areas where individuals make poor decisions *for themselves*, including in areas such as savings decision (Madrian and Shea, 2001; Thaler and Benartzi, 2004), health (Gine et al. 2008) and food choices (Dayan and Hillel, 2011; Downs et al., 2009). However, we have seen an increased interest in nudging in areas where individuals make decisions that involve negative externalities, and where hence individuals might make bad decisions *for the society*. By *prosocial* nudging, we mean nudging intended to reduce negative externalities from consumption, rather than dealing with internalities in the sense that people are nudged to make better decision for themselves.

A nudge is typically seen as a change in any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives (Thaler and Sunstein, 2008). This is true also for prosocial nudges. However, while a well-functioning conventional nudge “creates large benefits for those who make errors, while imposing little or no harm on those who are fully rational” (Camerer et al. 2003, 1212), the purpose of prosocial nudges is not to correct for individual mistakes. Instead it is to use people's cognitive and self-control limitations, as well as their moral convictions, in order to affect their behavior such that they will cause less negative externalities. Consequently, the rationale for nudging constitutes the fundamental difference between prosocial and conventional nudges.

While we are gaining increasing knowledge on when and how such prosocial nudges work, in the sense that individual behavior is affected in the intended direction, much less attention has been put on prosocial nudging as a policy instrument compared to other policy instruments, from a welfare point of view. The present paper in contrast analyzes the socially optimal use of various kinds of prosocial nudging as well as trade-offs between standard policies, such as environmental taxes, and nudges. We will also analyze when conventional policy instruments and nudges are perfect substitutes, and whether they can complement each other. Such optimal prosocial nudging is then analyzed in a number of different settings.

After a brief literature review on policy instruments with behavioral agents with a focus on prosocial nudging in Section 2, Section 3 outlines a simple model of prosocial nudging in an

economy with identical individuals and two goods, one clean and one dirty, and where there are no externalities. Two different kinds of costly nudges are introduced: cognitive and moral nudges. A cognitive nudge does not affect the utility function directly, but instead makes the dirty good to be perceived as less attractive relative to the clean one. A moral nudge, in contrast, affect the consumers' moral utility, e.g. the nudge makes it morally costly to consumer the dirty good relative to the clean one. We analyse two versions of ta moral nudge: encouraging and discouraging moral nudges. People are either encouraged to do the right thing, or discouraged to do the wrong thing. An encouraging moral nudge induces additional utility from consuming the clean good, whereas discouraging moral nudges induce disutility from consuming the dirty good. The optimal provision rules for both types of nudges depend strongly on whether or not it is possible to use a conventional Pigovian tax on the dirty good. It is never optimal to use costly cognitive nudges or discouraging moral nudges together with an optimal Pigovian tax on the dirty good, whereas it can then still be optimal to use an encouraging moral nudge.

In Section 4 we add initial biases with respect to the perceived price of the dirty good. We incorporate key features from the existing theoretical nudging literature focussing on externalities, such as Allcott et al. (2014), Allcott and Kessler (2015) and Farhi and Gabaix (2017), into the theory of optimal prosocial nudging. The optimal provision rules for the nudge then reflects both externalities and the initial biases, and they still critically depend on whether or not a conventional Pigovian taxes can be used.

Section 5 incorporates the often observed empirical pattern that people's prosociality, for example charitable contributions, has a relative dimension, such that people prefer to be more prosocial than others, *ceteris paribus*. Section 6 generalizes the model to the case of heterogeneous individuals, who differ with respect to income, initial bias and nudgeability (how much their choices are affected by a nudge). The model then naturally becomes more complex, but by assuming additively separable and quasi-linear preferences, we are able to obtain distinct and intuitive results that abstract from distributional and income effects in a convenient way. In addition to average externalities and measures of nudges and nudgeability it is demonstrated how various covariances become parts of the optimal policy rules. Section 7 provides some concluding remarks and suggests directions for future research.

2. Literature on Prosocial Nudging

There is an increasing number of empirical studies investigating if various types of nudges affect individual behavior, and in turn environmental quality; Carlsson et al. (2018) provides a recent overview on nudges related to environmental issues. Prominent examples are default options for the choice of energy contract (e.g. Pichert and Katsikopoulos, 2008), social norm messages to affect energy- and water use (e.g. Allcott, 2011; Ferraro et al., 2011), order effects for environmentally friendly food choices (Kurz, 2018), and changes in the physical environment to reduce food waste (Kallbekken and Saelen, 2013). Other empirical studies on prosocial nudging have investigated behavior such as organ donations (Johnson and Goldstein, 2003), tax compliance (Hallsworth et al., 2017), and charitable giving (Shang and Croson, 2009).

The theoretical literature is much more limited, but there is a small literature on optimal environmental taxation when consumers suffer from some sort of internality, i.e. some kind of irrationality from the individual perspective (Allcott et al., 2014), or when there are moral concerns/altruism (Johansson, 1997). Allcott et al. (2014) present a model where consumers choose between an energy efficient and an inefficient durable good. They allow for misoptimizing agents that do not correctly value how energy efficiency will affect their future utility. This is modelled as a valuation weight function in the utility function. Energy use result in a constant marginal damage. There is thus both an internality and an externality. In their model there is set of different consumer types with different valuation weight functions. An optimal tax consists of two additive separable parts: the marginal damage of the externality and the average marginal internality. When the marginal internality is positive, the optimal tax will be higher than the marginal damage. They then allow for two instruments, a tax and a subsidy. The reasons for why a policy maker would consider both these instruments are that they can affect different consumers with different types of internalities and that they can be more or less efficient with respect to reducing the internality. They show that the magnitude of the optimal tax depends on how efficient it addresses the misoptimization relative to how efficient the subsidy is addressing the same misoptimization. In Tsvetanov and Segerson (2013) it is shown that a Pigovian tax is not sufficient to achieve a first-best outcome if consumers have self-control problems. In their particular setting, consumers have to decide between durable-goods with different energy efficiency. Similarly, Heutel (2015) look at incentive-based policies and command-and-control policies addressing externalities when consumers are present biased. It is shown that a standard pigovian tax will never be optimal; instead, a tax, or a command-and-control policy, would have to correct the present bias as well.

Moreover, there is an even smaller literature on optimal nudging. Allcott and Kessler (2015), and Farhi and Gabaix (2017) are the papers closest to the present one. In Allcott and Kessler (2015) there is a population of heterogeneous consumers, who choose the amount of a numeraire good and energy use. They allow for a difference between experienced utility and decision utility, so that there is potential behavioral bias. In addition, there is a moral utility term, which depends on energy use. Energy use also results in a negative externality. The policy maker can implement a nudge at a cost per consumer. Their model illustrates that the impact of a nudge on consumer and social welfare critically depend on how the nudge affects the consumer. A nudge that reduces the behavioral bias would be beneficial for the consumer and for society as well, if this reduce the extent of the externality, as would an encouraging nudge that increases the moral subsidy of low energy use be. On the other hand, a nudge that raises the moral cost of energy use would be bad for the consumer, but potentially good for society.

Farhi and Gabaix (2017) focus on optimal taxation with behavioral agents, but also discuss optimal nudges. In a general setting where individuals maximize decision utility, given perceived prices, and where perceived prices can be changed with a costless nudge then the optimality condition for the tax is unchanged and the optimality condition for the nudge has four terms. The first is the effect on welfare due to reduced misoptimization, the second is direct effects of the nudge on utility. The third terms is the direct effect on tax revenue of changes in behavior, and the final terms capture indirect effects on tax revenue and welfare through the effect of the nudge on the externality. They also discuss another role of a nudge, which is if a tax has undesirable distributional effects. In particular, a nudge as opposed to a tax, allows for a correction of the internality without directly affecting the income of the individuals.

3. Optimal Prosocial Nudging without Initial Biases

We begin with a model where there are no internalities present for the individual. There are no misperceptions of prices, nor any differences between decision and experienced utility. This simple model will serve as a way to introduce the basic set-up, but also allows us to explore if a nudge should be used in such a setting as well. We assume that each of n identical individuals cares about own consumption of two goods, one clean good C and one dirty good D . The dirty good gives rise to a negative externality which depends on the total consumption of good D : $E = E(nD)$. The government can use a consumption tax on the dirty good t , a lump-sum tax (subsidy), T , and a costly nudge Z . The production technology is linear and the production

prices of both goods are normalized to unity whereas the production price of the nudge is given by Q ; we will denote the average production price of the nudge by $q = \frac{Q}{n}$. The resource constraint is $nY = nC + nD + QZ$ or $Y = C + D + qZ$, where Y is the exogenously given income level for each individual.

Thus, we assume that there is a variable cost associated with a nudge, but not with a tax. This might not always be the case. There might be nudges that are more or less without any implementation costs and nudges that only have a fixed cost. The best example is perhaps default nudges, where either a default option is implemented, or the existing default option is redesigned. However, many other nudges would require non-negligible fixed costs and/or variable costs. Yet, we will also discuss cases when the marginal cost of the nudge is zero ($q = 0$).

We distinguish between *cognitive* nudges and *moral* nudges. A cognitive prosocial nudge uses a person's bounded rationality or inattention to affect individual behavior. It includes aspects such as designing the default alternative, or affecting the salience of certain aspects of the choice architecture. A moral nudge, on the other hand, draws on people's social preferences, their desire for status, to follow norms and to have positive self-image.

3.1 Cognitive Nudges

We model a cognitive nudge with an effect of the nudge on the perceived price of the dirty good. Consider a quasi-concave utility function, the same for each individual, given by:

$$U = u(C, D, E), \quad (1)$$

where $u_C > 0$, $u_D > 0$ and $u_E < 0$. The nudge, Z , affects the perceived price of good D when making the consumption choice. Each individual maximizes utility subject to the perceived budget constraint, given by:

$$\frac{Y - T}{1 - \Omega(Z)} = C + D(1 + t + \eta Z), \quad (2)$$

instead of its actual budget constraint

$$Y - T = C + D(1 + t), \quad (3)$$

where $\eta \geq 0$ captures the nudgability of the individual (cf. Gabaix and Farhi, 2017). We denote the perceived (relative) price of good D , $(1 + t + \eta Z)$, by P , the derivative of the demand with

respect to the perceived price by D_p , and the derivative of the demand with respect to the nudge by D_Z , where $D_Z = \eta D_p$. Thus, we assume that the nudge can increase the perceived relative price of the dirty good. $\Omega(Z)$ then simply reflects the correspondingly changed budget, in order for both the perceived and actual budget to hold simultaneously. For (2) and (3) to hold simultaneously we have that $\Omega(Z) = \frac{D}{C + D(1 + t + \eta Z)} \eta Z > 0$, where we thus assume that the individual will perceive income as exogenous.

Throughout the paper, we will let a cognitive nudge affect the perceived price. Of course, we do not believe that all cognitive nudges affect the perceived prices; this is instead used as a simple way to let the nudges affect the choices made without directly affecting utility. Another way to model what we denote a cognitive nudge is to distinguish between decision and experienced utility (Farhi and Gabaix, 2017), where decision utility but not experienced utility would be affected by the nudge. The results would be very similar with such an approach. This type of model is also similar to how inattention to prices (see e.g. Chetty, 2015; Allcott and Taubinsky, 2015) and product attributes (see e.g. Bordalo et al., 2013) is typically modeled.

Given these assumptions, we can state our first proposition.

Proposition 1. *With an optimal tax, and where the individuals initially do not suffer from any biases there is no role for a cognitive nudge. The optimal tax on the dirty good is the conventional Pigovian tax:*

$$t = t^{Pigou} = -nMRS_{EC} E'(nD).$$

The intuition is straightforward. If there are two ways to affect the perceived price, and where the marginal cost of affecting the price is zero for one - the tax - and possibly positive for the other - the nudge - then it can never be welfare improving to use a nudge. Yet, if the nudge is free of costs as well, then it would be equally optimal to use a nudge instead of the tax.

If the tax is not set at the optimal level then it would, under most circumstances, be optimal to use a nudge. Naturally, we will focus our interpretation of the results on the situation when the tax is set lower than the optimal tax, but all results hold for both cases. The individual is still

making the same maximisation problem as above, but with the difference that the tax is not optimal. We then have the following proposition:

Proposition 2. *With a tax on good D that is not optimal, and where individuals initially do not suffer from any bias, a cognitive nudge will be optimal under special conditions. We have that:*

i) For $Z > 0$, the optimal nudge is given by:

$$Z = \frac{\frac{q}{D_Z} + t^{Pigou} - t}{\eta}, \text{ else } Z = 0.$$

ii) For any level of the cognitive nudge the optimal level of the tax is given by:

$$t = t^{Pigou} - \eta Z.$$

The optimal level of the nudge in the case of no tax and no cost of providing the nudge is

$$Z = \frac{t^{Pigou}}{\eta}. \quad (4)$$

The optimal level of the nudge is the marginal damage divided by the nudgeability parameter. The intuition is straightforward. Let us take the optimal tax first. The purpose of the optimal tax is to change the relative prices of the two goods so that the externality is internalized. Consider next a nudge without any tax and with no implementation costs. The optimal amount of a nudge would then be such that the *perceived* relative price between the dirty and the clean good would exactly mimic the *actual* relative price that a Pigovian tax would induce. From the individual perceived budget restriction (2) we see that a nudge such that $\eta Z = t^{Pigou}$ would change the perceived relative price equally much as a Pigovian tax would.

Similarly, in the case where a tax has internalized, say, 40% of the externality, and where there is still no costs associated with nudging per se, it follows that an optimal nudge would be given

as $Z = \frac{t^{Pigou} - t}{\eta} = \frac{t^{Pigou} - 0.4t^{Pigou}}{\eta} = 0.6 \frac{t^{Pigou}}{\eta}$. Thus, an optimal nudge should then be such that

it would cause an equally large relative price change as a tax that would internalize the remaining part of the externality would.

When we introduce a positive cost of nudging, it naturally becomes less attractive to nudge. The optimal level of nudging will now decrease with the price of the nudge in relation to how effective the nudge is in affecting the perceived price.

In our framework, proposition 2 implies that it is optimal to have a cognitive nudge at all if and only if $-\frac{q}{D_Z} < t^{Pigou} - t$.¹ Thus, because of the cost of providing the nudge, the optimal nudge will be smaller than the difference between the optimal tax and the actual tax scaled by the nudgability parameter. The left hand side of this expression is increasing in the marginal cost of the nudge, q , and decreasing in the effectiveness of the nudge, D_N . The optimal level of the nudge is higher the lower the tax is, and higher the lower η is. A lower η means that individuals are less affected by the nudge, and thus to achieve a given effect on behavior the nudge has to be stronger

3.2 Moral Nudges

So far, we have assumed that there are no direct effects on experienced utility of a nudge, beyond those effects caused by the changes in consumption choices. However, many nudges aim at affecting the experienced utility directly; in particular nudges that relate to norms. It could for example be nudges that gives information about norms or strengthen the impact of the norms. We will here assume that the nudge only affects a norm, which in turn will affect experienced utility, and hence decision utility, but not the perceived price. Following Levitt and List (2007) we let individuals care about perceived own morality, M , in addition to own consumption of the two goods and the externality:

$$U = u(C, D, E, M), \quad (5)$$

where $u_C > 0$, $u_D > 0$, $u_E < 0$ and $u_M > 0$ and where u is strictly quasi-concave and twice continuously differentiable. The perceived own morality is monotonically decreasing in own consumption of the dirty good D , and the strength of this norm is influenced by the nudge Z , such that: $M = M(D, Z)$. Each individual maximizes utility subject to the actual budget constraint, given by:

$$Y - T = C + D(1+t). \quad (6)$$

¹ Note that we have modelled the costs of nudging in a very simple way. If we would introduce fixed costs of nudging, the conditions for when it is optimal to nudge would naturally be more stringent.

We can think of different broad types of moral nudges, intended to reduce consumption of the dirty good D . One is an *Encouraging moral nudge*, which encourage people to do the right thing, i.e. consume less of the dirty good. Introducing or increasing such a nudge would mean that people get more utility *ceteris paribus*, i.e. for any given behaviour, such that $M_Z > 0$ (recall that people's utility increases in M). In addition, an encouraging moral nudge may reduce the consumption of the dirty good since the benefit of behaving in an appropriate manner is even larger. Another type of nudge is a *Discouraging moral nudge*, which discourage people from doing the wrong thing, which here means to discourage them from consuming the dirty good. Introducing or increasing such a nudge would mean that people get less utility *ceteris paribus*, i.e. for a given behaviour, such that $M_Z < 0$. In addition, a discouraging nudge may reduce the consumption of the dirty good since the cost of not behaving in the appropriate way has increased.

Another way to think about this is to think about one moral nudge but distinguish between different individuals with different levels of consumption of the dirty good. In particular, consider the following additively separable formulation:

$$U = u(C, D, E, M) = f(C, D, E) + M = f(C, D, E) + g(Z)h(\hat{D} - D) \quad (7)$$

where $g(Z) > 0$, $h(D) > 0$, $g'(Z) > 0$, $h'(D) < 0$ and where \hat{D} denotes the reference consumption level; moral utility is thus given by $g(Z)h(\hat{D} - D)$. First, assume that $\hat{D} > D$, where e.g. \hat{D} is the amount of the dirty good that the individual would have consumed without any morality. Here clearly people receive a positive utility effect, or an increased warm-glow effect, from an increased nudge without considering any consumption adjustment due to the nudge so $M_Z = g'(Z)h(\hat{D} - D) > 0$, thus the nudge would be an encouraging nudge. Consider next a discouraging nudge, which can be illustrated with the same functional form but where now instead $\hat{D} < D$, where a special case is that $\hat{D} = 0$ so that people simply receive a negative utility effect due to their consumption of the dirty good. Here $M_Z = -g'(Z)h(D) < 0$, i.e. the moral utility will decrease with an increased nudge *ceteris paribus*.

This functional form also points to a third case which is where in equilibrium $\hat{D} = D$. This means that $M_Z = 0$, such that utility on the margin is independent of the nudge, *ceteris paribus*. We will denote this case as a *Hedonically neutral moral nudge*. An increase of a hedonically

neutral moral nudge will neither make you feel better or worse about yourself than before, but it would make the moral issues more salient such that you would still buy less of the dirty good.

Despite the differences between cases, an increasing nudge will for each of these cases imply that the value of the dirty good for the consumers, as measured by the marginal rate of substitution between the dirty good and the clean good, will decrease with the nudge *ceteris paribus*. Indeed, it will decrease in exactly the same way in each of the cases, as follows

$$\left(\frac{\partial MRS_{DC}}{\partial Z} \right)_{c,D} = \frac{\partial \left(\frac{f_D - g(Z)h'}{f_C} \right)}{\partial Z} = -\frac{h'}{f_C} g'(Z) < 0 \quad (8)$$

For the optimal policy, it turns out to be important to distinguish between encouraging, discouraging and hedonically neutral moral nudges, as can be observed in the following proposition:

Proposition 3. *With an optimal tax, and where the individuals initially do not suffer from any biases,*

i) then for an encouraging moral nudge ($M_Z > 0$) an interior solution of the optimal moral nudge is given by:

$$MRS_{MC} M_Z = q, \text{ else } Z = 0.$$

ii) It is not optimal to provide a discouraging or hedonically neutral moral nudge ($M_Z \leq 0$).

iii) The optimal tax on the dirty good is the Pigovian tax: $t = t^{Pigou}$.

Thus, a moral nudge should only be used if it directly increases perceived moral utility, i.e. only encouraging nudges should potentially be used. Furthermore, the only role of the nudge is to directly affect moral utility, since the tax can deal with the externality *per se*. The nudge would thus function as a way of rewarding individuals that are doing something good, and the marginal utility of this reward should be equal to the marginal cost of the nudge. Do note, however, that the level of the tax with moral utility could be different from the level of the tax without moral utility, and the level of the optimal tax could vary with the level of the nudge as well.

Again, we next consider the optimal moral nudge when the tax for whatever reason is not optimal:

Proposition 4. *With a tax on D that is not optimal, and where the individuals initially do not suffer from any bias, then*

i) For $Z > 0$, the optimal moral nudge is given by:

$$-\frac{q}{D_Z} = t^{Pigou} - t - MRS_{MC} \frac{M_Z}{D_Z}, \text{ else } Z = 0$$

ii) For any level of the moral nudge the optimal level of the tax is given by:

$$t = -nMRS_{EC} E'(nD) = t^{Pigou}.$$

For simplicity, let us first set the tax on the dirty good to zero. For $Z > 0$, the optimal level of the nudge is then given by:

$$q = MRS_{MC} M_Z - t^{Pigou} D_Z. \quad ()$$

The left-hand-side of the expression for the optimal level of the nudge is the marginal cost of the nudge. The first term on the right-hand-side reflects the monetary value of the direct effect of the nudge on utility through effects on perceived morality (for a given consumption of the dirty good). The second term on the right-hand-side side reflects the marginal damage of the externality and the effectiveness of the moral nudge on the demand of the dirty good.

If the moral nudge on the margin is hedonically neutral ($M_Z = 0$), then the first term on the left-hand-side is zero. The optimal level of the nudge depends only on the marginal damage, the effectiveness of the nudge in terms of affecting the consumption of the dirty good and the marginal cost of the moral nudge. If the nudge is encouraging ($M_Z > 0$), then there is an additional welfare enhancing effect of the nudge, which implies, ceteris paribus, a higher level of the nudge compared with a hedonically neutral nudge. If the nudge is discouraging ($M_Z < 0$), then there is welfare reducing effect of the nudge in terms of moral utility and a welfare enhancing effect in terms of reduced level of the externality. This parallels the analysis in Allcott and Kessler (2015), who analyse an optimal nudge where there is both an internality and a moral utility, and show that the effect on the consumer (and the size of the nudge) critically depends on if the nudge increases or decreases moral utility. This also means that if the negative effect on moral utility of a discouraging nudge is sufficiently large, the optimal level of the moral nudge is zero.

It is important to point out that the above discussion rests on the assumption that the effect on the demand of the dirty good is similar for a nudge that affects guilt as for a nudge that has a positive effect on utility. The literature on norm messages related to water and energy use provides some empirical evidence about the plausibility of this assumption (see e.g. Allcott, 2011; Ferraro and Price, 2013). With these norm messages, low users receive a positive feedback on their use, while the high users receive a negative feedback on their use. This would suggest that the effect could be stronger for those receiving a negative effect on moral utility. However, it could also be that there is simply more room to improve for high users, or that they actually obtain a positive moral utility from the actual improvement in behaviour. The latter is not how we have modelled the effect on moral utility though.

Another interesting question is how to view moral utility from a social welfare perspective. So far, we have assumed that social planner considers moral utility when setting the optimal tax and nudge. As we saw, this results, among other things, that it is optimal for the government to implement encouraging moral nudges even with an optimal tax because of a resulting increase in moral utility. One might argue that this type of welfare effects should not be taken into consideration when designing the policies, although one could still rely on the mechanism that the moral nudge provides. The condition for the optimal level of the nudge is then actually equivalent to the case of a hedonically neutral nudge.

4. Optimal Prosocial Nudging with an Initial Bias

So far we have assumed that individuals have no behavioral biases, and found that the role of nudges is rather limited. The traditional role of nudges is when there are behavioral biases, and the question is of course if the presence of a negative externality affects the role of a nudge. As before we assume that the government can use a consumption tax on the dirty good t , a lump-sum tax (or subsidy if negative), and a costly nudge Z in order to affect the perceived price by the consumer when making the consumption choice. In addition, we assume the consumer faces an initial bias, b , of the perception of the price of the dirty good, D . This bias can be either positive or negative. We will primarily focus on the case where there is an underestimation of the real price. It could for example be inattention to the price or certain aspects of the price (Allcott, 2013; Tanjim and Morgan, 2006; Chetty et al. 2009), or biased beliefs (Allcott, 2013, Allcott and Wozny, 2014). It is important to point out that there is nothing inherent in that there would be a bias in the perception of the price of dirty goods only, and not on the price of clean goods. However, we will restrict our attention to this case.

4.1 Cognitive Nudge

Let us again begin with the case where there is no moral utility, and the nudge affects the perceived price of the dirty good. Each individual maximize utility subject to the perceived budget constraint, given by:

$$\frac{Y-T}{1-\Omega(Z)} = C + D(1+t-b+\eta Z), \quad (9)$$

where b reflects the initial bias of the perceived price on the dirty good D . We assume that the nudge will increase the perceived relative price of the dirty good, while $\Omega(Z)$ reflects the correspondingly changed budget such that the perceived and actual - still given by (3) - budget constraints hold simultaneously, implying that $\Omega(Z) = \frac{D}{C + D(1+t-b+\eta Z)}(\eta Z - b)$. The production technology and the resource constraint are the same as before. Given that the individuals maximises the utility function in (1), and the above perceived budget constraint we have the following proposition.

Proposition 5. *With an initial behavioural bias, and a costly cognitive nudge we have that:*

- i) *If the tax is optimal there is no role for a cognitive nudge: The optimal tax on the dirty good is given by a modified Pigovian tax: $t = b - nMRS_{EC} E'(nD) = b + t^{Pigou}$.*
- ii) *If the tax is not optimal then the optimal cognitive nudge is given by:*

$$Z = \frac{\frac{q}{D_Z} + b + t^{Pigou} - t}{\eta}$$

- iii) *For any level of the cognitive nudge the optimal tax is given by:*

$$t = t^{Pigou} + b - \eta Z$$

The optimal tax is similar to what Farhi and Gabaix (2017) derive: the optimal Pigovian tax is modified by a term due to a behavioural bias such as a misperception of prices. The larger the initial bias is, the larger the correction of the original level of the Pigovian tax needs to be. Again, we find that if an optimal tax can be implemented, there is no role for a nudge, despite a behavioural bias. Do note that we have assumed that individuals perceive the tax without bias. If the individual would only perceive a fraction of the tax, then the optimal tax would have to

be corrected due to this as well (Farhi and Gabaix, 2017). However, it would not change the conclusion that the tax is preferred.

Without a tax and with zero cost of providing the nudge, the condition for the optimal level of the nudge simplifies to:

$$Z = \frac{b + t^{Pigou}}{\eta}. \quad (10)$$

The expression for the optimal nudge with a non-optimal tax is similar to the expression derived without a behavioural bias (proposition 2). The difference is that the optimal nudge also depends on the initial bias. For the case when $b > 0$, i.e. that the consumer initially does not perceive the true price of the dirty good, it is more likely that it is optimal to use a nudge. Furthermore, the level of the nudge is increasing in the absolute magnitude of this initial bias.

4.2 Moral Nudge

Next, we reconsider the case of a moral nudge when allowing for an initial bias as well. We then have the following proposition:

Proposition 6. *With an initial behavioural bias, and a costly moral nudge we have that:*

i) *If the tax is optimal, an interior solution of optimal moral nudging is given by the same choice rule as without initial behavioural biases (proposition 3). The optimal tax on the dirty good is given by the same modified Pigovian tax as for the case where nudging does not affect experienced utility directly (proposition 3).*

ii) *If the tax is not optimal then, for $Z > 0$, the optimal moral nudge is given by:*

$$-\frac{q}{D_Z} = t^{Pigou} - t + b - MRS_{MC} \frac{M_Z}{D_Z} \text{ else } Z = 0.$$

iii) *For any level of the nudge, the optimal tax is given by*

$$t = t^{Pigou} + b.$$

These results follows naturally from the discussion around propositions 3 and 5. Again, the optimal nudge, when the tax is not optimal, depends on the sign and size of the initial behavioural bias. With $b > 0$, a higher bias would imply a higher level of the nudge. The direct effect on moral utility is of course still relevant as well. With $M_Z > 0$ the level of the nudge would tend to be higher, since there is a direct positive effect on moral utility of the nudge. The

expression of the optimal level of the tax for a given level of the nudge is the same as the expression for the optimal tax, but the level of the tax is likely to be different.

5. Optimal Prosocial Moral Nudging with Social Status

We now consider a model with moral nudges and social status concerns. People care about absolute consumption as well as relative consumption of both goods, such that they prefer to consume more than others, *ceteris paribus*. In addition, they care about morality, which in turn depends on their own absolute as well as relative consumption of the dirty good, as well as the moral nudge (which can here too be either encouraging or discouraging). While the extension of social comparisons naturally makes the model more complex, it is realistic both with respect to consumption or income (e.g. Johansson-Stenman et al. 2002, Alpizar et al. 2005, Clark et al. 2008) and prosociality such as giving to charity (e.g. Glazer and Konrad, 1996; Harbaugh, 1998a,b). Thus, we will model a case where the individual has positional concerns for both the clean good, C , and the dirty good, D , and where relative consumption of D matters for morality, as follows:

$$U = u(C, C - \bar{C}, D, D - \bar{D}, E, M(D, D - \bar{D}, Z)), \quad (11)$$

where a bar denotes the population average of the variable. The assumption that status depends on the average is of course a simplification, but an innocuous one here when people are identical. Following for example Aronsson and Johansson-Stenman (2008), define the degree of positionality of good C as:

$$\alpha^C = \frac{u_{C-\bar{C}}}{u_C + u_{C-\bar{C}}}. \quad (12)$$

We can interpret α^C as representing the fraction of the overall utility increase from the last dollar spent on C that is due to the increased relative consumption. $\alpha^C = 0$ represents the conventional case where only absolute consumption matters, whereas $\alpha^C = 1$ represents the case where, on the margin, only relative consumption of C matters to the individual. Let α^D denote the corresponding degree of positionality for good D ; for simplicity we assume the same degree of positionality for the two goods: $\alpha^C = \alpha^D = \alpha$. We define the degree of moral positionality as:

$$\beta = \frac{M_{D-\bar{D}}}{M_D + M_{D-\bar{D}}}. \quad (13)$$

The interpretation is similar as for the consumption of the two goods: β represents the fraction of the overall moral utility increase from the last dollar reduction spent on the dirty good D that

is due to an improved relative consumption of D . $\beta = 0$ represents the case where morality only depends on own absolute consumption of the dirty good D , whereas $\beta = 1$ represents the case where only relative consumption of the dirty good D matters for morality.

Before presenting the results, let us consider two different definitions of the marginal rate of substitution, depending on whether the consumption and morality of others are held fixed or not. These two definitions will be important for the interpretation of the results regarding the optimal levels of the nudge and the tax. The conventional marginal rate of substitution between environmental quality and private consumption ceteris paribus, i.e. holding everything else in the model fixed, including the consumption of others, is:

$$MRS_{EC} = \frac{u_E}{u_C + u_{C-\bar{c}}}, \quad (14)$$

Analogous to the definition in Aronsson and Johansson-Stenman (2008) for public goods, let us next define

$$CMRS_{EC} = \frac{u_E}{u_C}. \quad (15)$$

as the marginal rate of substitution between environmental quality and private consumption when consumption is changed equally much for all individuals in society. Correspondingly, let us define:

$$MRS_{MC} = \frac{u_M}{u_C + u_{C-\bar{c}}}, \text{ and} \quad (16)$$

$$CMRS_{MC} = \frac{u_M}{u_C}. \quad (17)$$

If individuals maximises the utility function in (11) under the budget constraint $Y - T = C + D(1+t)$, we have the following proposition regarding optimal nudging.

Proposition 7. *With status concerns and an optimal tax,*

i) an interior solution of an optimal encouraging moral nudge ($M_Z > 0$), is given by

$$q = CMRS_{MC} M_Z = \frac{MRS_{MC} M_Z}{1-\alpha}.$$

ii) It is never optimal to provide a discouraging moral nudge ($M_Z < 0$).

iv) *The optimal tax on the dirty good is:*

$$t = t^{\text{Pigou}C} + \frac{\beta - \alpha}{1 - \beta} CMRS_{MC} M_D,$$

or

$$t = \frac{t^{\text{Pigou}}}{1 - \alpha} + \frac{\beta - \alpha}{(1 - \alpha)(1 - \beta)} MRS_{MC} M_D$$

The conditions for the optimal level of the nudge can be expressed in almost the same way as in the model with moral concerns without status concerns: the marginal cost of the nudge should be equal to the marginal benefit of the nudge in terms of improved moral utility, where the only difference is that MRS_{MC} is replaced by $CMRS_{MC}$. The latter measures the individual marginal willingness to pay for improved morality when all individuals in the economy will have to pay for the moral improvement. As such, it can be seen as a collective decision regarding how much people would be willing to pay for improved morality if all people would make the choice *together*. Note that $CMRS_{MC}$ in principle reflects the marginal willingness to pay measures that would be revealed based on a so-called referendum approach in the stated preference literature, i.e. an approach where people are asked to vote yes or no to a hypothetical referendum for a certain (e.g. environmental) improvement where all citizens would have to pay for the improvement. MRS_{MC} , in contrast, reflects how much an individual would be willing to pay for the moral improvement *in isolation*. Such a marginal payment would have three utility consequences for the individual: increased morality and decreased consumption, and improved relative consumption (since the consumption of others would not decrease here); therefore $MRS_{MC} < CMRS_{MC}$. For this reason, MRS_{MC} would also reflect a desire not to fall behind the other individuals in terms of consumption. To correct for the latter effect, or rather the externalities associated with the fact that the individuals take such in the optimal choice rule MRS_{MC} is divided by $1 - \alpha$.

However, the expression for the optimal tax is different from the case with moral utility without status concerns. The reason is that the concern for status generates an additional externality. Whether the optimal tax is lower or higher than the standard Pigovian tax depends on a number of factors. First, the standard term should be scaled because of the concern for status in the consumption goods, α , similar as for the optimal nudge rule. In the first expression for the

optimal tax we denote the corresponding Pigovian tax by t^{PigouC} ; this is thus equal to $\frac{t^{Pigou}}{1-\alpha}$.

Second, if the marginal degree positionality is stronger for the moral good, β , than for the consumption goods, α , then the optimal tax is also smaller than the standard Pigvovian tax, since $M_D < 0$. Thus, if the concern for status in the moral utility is sufficiently strong, then the optimal tax will be reduced.

The conditions for the optimal level of a nudge if a tax is not set at the optimal level is given in Proposition 8.

Proposition 8. *With status concerns and a non-optimal tax, an interior solution of the optimal moral nudge is given by:*

$$-\frac{q}{D_Z} = t^{Pigou} - t + (\beta - \alpha)CMRS_{MC}M_D - CMRS_{MC}\frac{M_Z}{D_Z}$$

For any level of the nudge, the optimal tax is still given by the same choice rule in proposition 7.

The condition for the optimal level of the nudge is now different from a model without social status; the reason is the same as for the optimal tax, the concern for status imply additional externalities that would need to be taken into consideration for the optimal level of the nudge.

6. Generalization to a Heterogeneous Population

We will now generalize our analysis to a setting with heterogeneous individuals. We consider the same models as before, where the government can use a consumption tax, t , on the dirty good, individual-specific lump-sum taxes (or subsidies if negative), and a costly nudge Z .

6.1 Cognitive nudge

Again, the cognitive nudge affects the perceived price by the consumer when making the consumption choices. We assume the following utility function:

$$U^i = C^i + g(D^i, E), \quad (18)$$

where $g_D > 0$ and $g_E < 0$; g is strictly concave and twice continuously differentiable in D and E . The additive separability and quasi-linearity assumptions are made for analytical simplicity,

and in order to be able to abstract from distributional and income effects in a convenient way. Each individual maximizes utility subject to the perceived budget constraint, given by:

$$\frac{Y^i - T^i}{1 - \Omega^i(Z)} = C^i + D^i(1 + t - b^i + \eta^i Z). \quad (19)$$

For this to hold simultaneously with each individual's actual budget constraint, given by $Y^i - T^i = C^i + D^i(1 + t)$ we have that $\Omega^i(Z) = \frac{D^i}{C^i + D^i(1 + t - b^i + \eta^i Z)}(\eta^i Z - b^i)$. Note that we

allow for heterogeneity in the initial bias and nudgeability as well. We will use the following

short notations for the normalized covariances: $\delta_{\eta, D_p} \equiv \text{cov}\left(\frac{\eta}{\eta}, \frac{D_p}{D_p}\right)$, $\delta_{b, D_p} \equiv \text{cov}\left(\frac{b}{b}, \frac{D_p}{D_p}\right)$,

$\delta_{\eta, D_z} \equiv \text{cov}\left(\frac{\eta}{\eta}, \frac{D_z}{D_z}\right)$, and $\delta_{b, D_z} \equiv \text{cov}\left(\frac{b}{b}, \frac{D_z}{D_z}\right)$. Note that these are the normalized covariances,

which will have implications for the sign of the covariance. For example, if $\delta_{\eta, D_p} > 0$, then individuals with a high nudgeability also have a demand of the dirty good that is more sensitive to a change in perceived price.

The externality is given by:

$$E = E(D^1 + \dots + D^n) = E\left(\sum_i D^i\right). \quad (20)$$

The production technology is still linear and the same as the one with identical individuals; although the individuals and their income and consumption now differ. Hence:

$$\sum_{i=1}^n Y^i = \sum_{i=1}^n C^i + D^i + QZ. \quad (21)$$

The government can finance the nudge by individual-specific lump-sum taxes as well as a tax on the dirty good (the same tax rate for all), such that each individual faces the actual budget constraint:

$$Y^i - T^i = C^i + (1 + t)D^i. \quad (22)$$

The aim of the government is to choose the tax level, t , the nudge, Z and the individual lump-sum taxes/transfers, to maximize a utilitarian social welfare function:

$$W = \sum_{i=1}^n U^i. \quad (23)$$

We then have the following proposition regarding the optimal nudge and tax:

Proposition 9. *With a heterogeneous population with respect to initial bias and nudgeability, and an optimal tax, then:*

- i) *There may be a role for a cognitive nudge. For $Z > 0$ the optimal level of the cognitive nudge is:*

$$N = \frac{\frac{q}{D_Z} + \bar{b}(\delta_{b,D_Z} - \delta_{b,D_p})}{\eta(\delta_{\eta,D_Z} - \delta_{\eta,D_p})}$$

- ii) *The optimal tax is given by:*

$$t = t^{Pigou} + \bar{b} \left((1 + \delta_{b,D_p}) + \frac{\delta_{b,D_Z} - \delta_{b,D_p}}{\delta_{\eta,D_Z} - \delta_{\eta,D_p}} (1 + \delta_{\eta,D_p}) \right) - \frac{q}{D_Z} \frac{1 + \delta_{\eta,D_p}}{\delta_{\eta,D_Z} - \delta_{\eta,D_p}} \text{ if } Z > 0$$

$$t = t^{Pigou} + \bar{b} (1 + \delta_{b,D_p}) \text{ if } Z = 0$$

Once we introduce heterogeneity, there are some instances when a nudge is socially optimally even when an optimal tax can be implemented. Before proceeding, do note that a nudge would only be optimal under a number of conditions. Apart from that the cost of the nudge cannot be too high relative to the extent of the bias and the nudgeability, there are also some conditions on the covariance terms.² In order to facilitate interpretation let us begin with a more restrictive case. Let us assume that the two covariance terms, δ_{b,D_p} and δ_{η,D_p} , are zero. This means that the distribution of the initial bias and the nudgeability is completely independent of the distribution of the price sensitivity of the dirty good. In other words, the effect of the tax on individuals demand is completely uncorrelated with the initial bias or the nudgeability. In that case, the condition for the optimal nudge reduces to:

$$N = \frac{\frac{q}{D_Z} + \bar{b}(\delta_{b,D_Z})}{\eta\delta_{\eta,D_Z}} = \frac{\frac{q}{D_Z} + \bar{b}(\delta_{b,D_Z})}{\eta \text{ var}(\eta)}. \quad (24)$$

Let us focus on the case where $\bar{b} > 0$, i.e. the average bias is an underestimation of the price of the dirty good. A nudge would only be optimal iff $\frac{q}{D_Z} + \bar{b}(\delta_{b,D_Z}) > 0$. Given that it is optimal to use a nudge, the level of the nudge would be increasing in the extent of the bias, \bar{b} , and decreasing in the nudgeability, $\bar{\eta}$. The term δ_{b,D_Z} is the normalized covariance between the

² Either $(\delta_{b,D_Z} - \delta_{b,D_p}) > 0$ and $(\delta_{\eta,D_Z} - \delta_{\eta,D_p}) > 0$, or $(\delta_{b,D_Z} - \delta_{b,D_p}) < 0$ and $(\delta_{\eta,D_Z} - \delta_{\eta,D_p}) < 0$.

initial bias and the responsiveness of the demand of the dirty good to the nudge. A positive covariance means that subjects with a high initial bias are more responsive to the nudge. Thus with a positive covariance, it will be more beneficial for society to use a nudge compared with a tax. Think of an extreme situation where only some individuals underestimate the price of the dirty good. An optimal tax would then be adjusted upwards for all individuals based on the average underestimation. Now, if those that underestimate the price are highly nudgeable while others are not affected by the nudge, then a nudge might be preferred over a tax because it would better target individuals with a biased perception. Finally, a higher variance in nudgeability implies a lower level of the optimal nudge. With a higher variance, there is an additional cost of a nudge in terms of the absence of an effect compared with a tax.

Let us now return to the more general case in proposition 9. In the numerator, we then also have the normalized covariance between the bias and the price sensitivity of the demand of the dirty good. A positive covariance means that individuals with a high bias are more sensitive to changes in the price of the good (and hence the tax), and therefore the optimal level of the nudge will be lower with a positive covariance. In the denominator, we have two covariance terms for nudgeability. If δ_{η, D_z} is positive, then the nudge is effective for those that are nudgeable. This implies that the nudge does not have to be that large to obtain a given effect. If δ_{η, D_p} is positive this means that those that are nudgeable are also more responsive to the price, and therefore the level of the nudge will be higher.

Let us now look at the case with a non-optimal tax:

Proposition 10. *With a heterogeneous population and a non-optimal tax then:*

i) *The optimal level of the cognitive nudge is given by*

$$Z = \frac{\frac{q}{D_z} + t^{Pigou} - t + \bar{b}(1 + \delta_{b, D_z})}{\bar{\eta}(1 + \delta_{\eta, D_z})}$$

ii) *The optimal tax on the dirty good for a given level of the nudge is given by:*

$$t = t^{Pigou} + \bar{b}(1 + \delta_{b, D_p}) - Z\bar{\eta}(1 + \delta_{\eta, D_p})$$

The optimal level of the nudge without a tax and with no cost of providing the nudge simplifies to

$$Z = \frac{t^{Pigou} + \bar{b}(1 + \delta_{b,D_Z})}{\bar{\eta}(1 + \delta_{\eta,D_Z})}. \quad (25)$$

The level of the nudge is increasing in \bar{b} (with a high average initial bias the nudge would have to be larger) and decreasing in $\bar{\eta}$ (if the nudge on average is effective then the nudge does not have to be that large). Furthermore, if the covariance term, δ_{b,D_Z} is positive, this means that those with a high initial bias are more responsive to the nudge, and therefore the level of the nudge will be higher ceteris paribus. Finally, if the covariance term δ_{η,D_Z} is positive, those that are nudgeable are also more responsive to the nudge, which means that the level of the nudge will be smaller.

6.2 Moral nudge

Extending the case of a moral nudge with heterogeneity, we assume the following utility function:

$$U^i = C^i + g^i(D^i, E, M^i). \quad (26)$$

If individuals maximise this utility function subject to the same perceived budget constraint as before, we then have the following proposition regarding the optimal nudge and tax.

Proposition 11. *Consider a model with a heterogeneous population and initial behavioural biases in a model where nudging affects a moral norm, and where externalities can be taxed optimally. Then:*

i) For $Z > 0$, the optimal level of the nudge is given by the condition

$$-\frac{q}{D_Z} = (\delta_{b,D_Z} - \delta_{b,D_p})\bar{b} + \frac{\overline{MRS_{MC}M_Z}}{D_Z}$$

ii) The optimal tax on the dirty good is given by

$$t = t^{Pigou} - \bar{b}(1 + \delta_{b,D_p})$$

As before, the optimal level of the nudge is decreasing in the marginal cost of the nudge, and the moral nudge should only be used there is a positive effect on moral utility of the nudge (an encouraging nudge). Moreover it will depend on the extent of the bias and the two normalized

covariance terms δ_{b,D_z} and δ_{b,D_p} . The level of the nudge will be higher if there is a positive covariance between the initial bias and the responsiveness of the demand of the dirty good to the nudge, and lower if there is a positive covariance between the initial bias and the responsiveness of the demand of the dirty good with respect to changes in price.

Similarly, we have the condition for the optimal level of the nudge when the tax is not optimal.

Proposition 12. *Consider a model with a heterogeneous population and initial behavioural biases in a model where nudging affects a moral norm, and where the tax is not optimal then:*

i) *The optimal level of the nudge is given by the condition*

$$-\frac{q}{D_z} = t^{Pigou} - t + \bar{b}(1 + \delta_{b,D_z}) + \frac{\overline{MRS_{MC}M_Z}}{D_z}$$

ii) *The optimal tax on the dirty good when nudging is not optimal is given by*

$$t = t^{Pigou} + \bar{b}(1 + \delta_{b,D_p})$$

These expressions are similar to the ones we have seen without moral utility, the difference is of course the additional term on the right hand side for the optimal level of the nudge. This term, in turn, is the same as in the homogenous case in proposition 6.

6.3 Moral nudge when there is status concerns

Finally, we extend the model with heterogeneity to allow for status concerns. We assume the following utility function:

$$U^i = C^i + g^i(C^i - \bar{C}, D^i, D^i - \bar{D}, E, M^i(D^i, D^i - \bar{D}, Z)), \quad (27)$$

where a bar denotes average in the population of the variable. Define the degree of positionality of good C and D for individual i as:

$$\alpha_C^i = \frac{g_{C^i - \bar{C}}^i}{1 + g_{C^i - \bar{C}}^i} \quad \text{and} \quad \alpha_D^i = \frac{g_{D^i - \bar{D}}^i}{g_{D^i}^i + g_{D^i - \bar{D}}^i}, \quad (28)$$

and the degree of moral positionality as

$$\beta^i = \frac{M_{D^i - \bar{D}}^i}{M_{D^i}^i + M_{D^i - \bar{D}}^i}. \quad (29)$$

As in the homogenous case we define:

$$MRS_{EC}^i = \frac{g_E^i}{1 + g_{C^i-\bar{C}}^i}, \quad (30)$$

as the conventional marginal rate of substitution for an individual between environmental quality and private consumption ceteris paribus, i.e. holding everything else in the model fixed, including the consumption of others. Furthermore, we define:

$$CMRS_{EC}^i = g_E^i, \quad (31)$$

as the marginal rate of substitution for an individual between environmental quality and private consumption when consumption is changed for all individuals in society. Correspondingly, define:

$$MRS_{MC}^i = \frac{g_M^i}{1 + g_{C^i-\bar{C}}^i} \text{ and } CMRS_{MC}^i = g_M^i. \quad (32)$$

We then have the following proposition:

Proposition 13. *Consider a model with a heterogeneous population, with status concerns and initial behavioural bias, and where externalities can be taxed optimally. Then,*

- i) *The optimal level of a moral nudge (if it is positive), given optimal taxation of the dirty good, is given by*

$$-\frac{q}{D_Z} = \frac{\bar{b}}{1-\alpha} (\delta_{D_p,b} - \delta_{D_Z,b}) - \frac{\overline{CMRS_{MC}M_Z}}{D_Z}$$

- ii) *The optimal tax is given by*

$$t = t^{PigouC} - \bar{b} \left(1 + \frac{\delta_{D_p,b}}{1-\alpha} \right) - \left(\frac{\beta - \alpha}{1-\beta} \right) \overline{CMRS_{MC}M_D} \left(1 + \delta_{\frac{\beta-\alpha}{1-\beta}, CMRS_{MC}M_D} \right)$$

For the level of the nudge, the main difference from the homogenous case is the first part of the left-hand side of the condition. What determines the sign of this is the relative size of the two normalized covariance terms. The first, $\delta_{D_p,b}$, is the normalized covariance between the bias and the responsiveness to the price. If this normalized, covariance is positive it means that those that have a high initial bias are also more responsive to the price. The second, $\delta_{D_Z,b}$, is the normalized covariance between the bias and the responsiveness to the nudge. If this normalized, covariance is positive it means that those that have a high initial bias are also more responsive to the nudge. Thus, similar to the case of a cognitive nudge, a moral nudge might be optimal to use even when an optimal tax can be used because it might better target individuals that have a biased perception of prices.

Finally, when the tax is not optimal, we have the following proposition:

Proposition 14 Consider a model with a heterogeneous population, with status concerns and initial behavioural bias, then,

- (i) The optimal level of a moral nudge when the dirty good cannot be taxed optimally is given by

$$-\frac{q}{D_Z} = t^{PigouC} - t - \bar{b} \left(1 + \frac{\delta_{D_z, b}}{1 - \alpha} \right) - \frac{\overline{CMRS}_{MC} M_Z}{D_Z} + \left(\frac{\beta - \alpha}{1 - \beta} \right) \overline{CMRS}_{MC} M_D \left(1 + \delta_{\frac{\beta - \alpha}{1 - \beta}, CMRS_{MC} M_D} \right)$$

- (ii) The optimal tax on the dirty good for any level of the nudge is given by

$$t = t^{PigouC} - \bar{b} \left(1 + \frac{\delta_{D_p, b}}{1 - \alpha} \right) - \left(\frac{\beta - \alpha}{1 - \beta} \right) \overline{CMRS}_{MC} M_D \left(1 + \delta_{\frac{\beta - \alpha}{1 - \beta}, CMRS_{MC} M_D} \right)$$

7. Conclusions

As far as we know, this is the first paper that has theoretically analyzed socially optimal levels of nudges in a prosocial context, such as dealing with environmental consequences, in a world where we to a varying extent could use a conventional Pigovian taxes to combat externalities. We find that under general settings, nudges would only play a limited role as long as optimal taxes can be implemented and the cost implementing a tax is lower than the cost of implementing a nudge. Moral nudges where people's choices are affected by strengthening the consumers' moral norms tend to maintain a larger role also when the consumption tax is optimal compared to purely cognitive nudges. By adding initial biases and internalities into our model we are able to incorporate important mechanisms from the existing theoretical nudging literature focussing on internalities with the prosocial nudging framework. Naturally, the optimal provision rules will then, in addition to the externality, also reflect the initial biases. We then show that relative concerns with respect to prosociality will also have important consequences for the social choice rule. The heterogeneous setting, finally, makes the model both more complex and realistic. In particular, this set of models points to cases where nudges would be optimal even when a first-best tax can be implemented. The reason why a nudge could be preferred is that it could better target individuals who for example suffers from an initial bias

than what a tax could do. More research is clearly needed, including dealing with non-atmospheric externalities and a simultaneous redistributive perspective.

Acknowledgments

Financial support from the Swedish EPA and the Center for Collective Action Research (CeCAR) is gratefully acknowledged. We have received valuable comments from seminar participants at the Swedish EPA, University of Gothenburg, University of Stirling, University of Toulouse, as well as participants at the World Conference of Environmental and Resource Economists in Gothenburg 2018.

References

- Allcott, H. 2011. Social norms and energy conservation. *Journal of Public Economics* 95(9): 1082-1095.
- Allcott, H. 2013. The welfare effects of misperceived product costs: Data and calibrations from the automobile market, *American Economic Journal: Economic Policy* 5, 30-66.
- Allcott, H., S. Mullainathan and D. Taubinsky (2014), Energy policy with externalities and internalities, *Journal of Public Economics* 112, 72-88.
- Allcott, H. and J. Kessler (2015), The welfare effects of nudges: A case study of energy use social comparisons, NBER Working Paper 21671.
- Allcott, H. and D. Taubinsky (2015) Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review* 105, 2501-2538.
- Allcott, H. and Wozny, N., Gasoline prices, fuel economy, and the energy paradox, *Review of Economics and Statistics*, 2015, 96, s. 779–795.
- Alpizar, F., F. Carlsson, and O. Johannson-Stenman (2005) How Much Do We Care about Absolute versus Relative Income and Consumption? *Journal of Economic Behavior and Organization* 56, 405–21.
- Aronsson, T. and O. Johannson-Stenman (2008), When the Joneses' consumption hurts: Optimal public good provision and nonlinear income taxation, *Journal of Public Economics* 92, 986-997.
- Bordalo, P., N. Gennaioli, S. Schleifer (2013) Saliency and consumer choice. *Journal of Political Economy* 121, 803-843.

- Camerer C, Issacharoff S, Loewenstein G, O'Donoghue T, Rabin M. 2003. Regulation for conservatives: behavioral economics and the case for “asymmetric paternalism.” *University of Pennsylvania Law Review*, 151:1211–54.
- Carlsson, F., C. Gravert, O. Johansson-Stenman and V. Kurz 2018. Nudging as an Environmental Policy Instrument. Working paper, department of Economics, University of Gothenburg.
- Chetty, R., Looney, A., and Kroft, K., 2009. Salience and Taxation: Theory and Evidence. *American Economic Review*, 99 (4), 1145–77.
- Chetty, R. (2015), Behavioral Economics and Public policy: A pragmatic perspective. *American Economic Review* 105, 1-33.
- Clark, A. E., P. Frijters and M. A. Shields (2008) Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature* 46, 95-144.
- Dayan, E., & Bar-Hillel, M. (2011). Nudge to nobesity II: Menu positions influence food orders. *Judgment and Decision Making*, 6(4), 333-342.
- Downs, J., G. Loewenstein, and J. Wisdom 2009. Strategies for Promoting Healthier Food Choices. *American Economic Review* 99 (2): 159–64.
- Farhi, E. and X. Gabaix (2017). Optimal taxation with behavioral agents.
- Ferraro, P., Miranda, J. and M. Price. 2011. The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review* 101(3): 318-322.
- Ferraro, P., and M. Price. 2013. Using non-pecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *The Review of Economics and Statistics* 95(1): 64-73.
- Giné, X., Karlan, D., & Zinman, J. (2010). Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4), 213-35.
- Glazer, A., K. Konrad, (1996) A Signaling Explanation for Charity. *American Economic Review* 86, 1019–1028.
- Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, 14-31.
- Harbaugh, W., (1998a) What do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow. *Journal of Public Economics* 67, 269–284.

- Harbaugh, W., (1998b) The Prestige Motive for Making Charitable Transfers. *American Economic Review* 88, 277–282.
- Heutel, G. (2015) Optimal policy instruments for externality-producing durable goods under present bias, *Journal of Environmental Economics and Management* 72, 54-70.
- Johansson, O. (1996) Optimal Pigovian taxes under altruism, *Land Economics* 73, 297-308.
- Johansson-Stenman, O., F. Carlsson, D. Daruvala, (2002) Measuring Future Grandparents' Preferences for Equality and Relative Standing. *Economic Journal* 112, 362–383.
- Johnson, E. J., Goldstein, D. (2003). Do Defaults Save Lives? *Science* 302 (5649): 1338–1339.
- Kalbekken and Saelen, H. (2011), ‘Nudging’ hotel guests to reduce food waste as a win–win environmental measure, *Economics Letters* 199, 325-327
- Kurz, V. (2018). Nudging to reduce meat consumption: Immediate and persistent effects of an intervention at a university restaurant. *Journal of Environmental Economics and Management*, 90, 317-341.
- Madrian, B. and D. Shea 2001. The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior, *The Quarterly Journal of Economics* 116, 1149-1187
- Pichert, D. and K. Katsikopoulos, 2008. Green defaults: Information presentation and pro-environmental behavior, *Journal of Environmental Psychology* 28, 63-73.
- Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *Economic Journal*, 119(540), 1422-1439.
- Tanjim, H. och Morgan, J., Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay. *Advances in Economic Analysis and Policy*, 2006.
- Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1), S164-S187.
- Thaler, R. H., and Sunstein, C. R. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. New York: Penguin Books.
- Tsvetanov, T. and C. Segerson (2013) Re-evaluating the role of energy efficiency standards: A behavioral economics approach, *Journal of Environmental Economics and Management* 66, 347-363.

Appendix

Proof of Proposition 1

The individual optimum condition implies that

$$MRS_{DC} = \frac{u_D}{u_C} = 1 + t + \eta Z = P. \quad (A1)$$

The social optimum, where the externality is taken into account, implies that

$$MRS_{DC} + n \frac{u_E E'(nD)}{u_C} = 1. \quad (A2)$$

Combining the private and the social optimum conditions, (A1) and (A2), we have

$$1 + t + \eta Z = 1 - n \frac{u_E E'(nD)}{u_C} = 1 - n MRS_{EC} E'(nD). \quad (A3)$$

It is clear that this expression is fulfilled with $t = t^{Pigou} = -n MRS_{EC} E'(nD)$ and $Z = 0$. It would also be fulfilled with other combinations where $Z > 0$, but that would be associated with a resource cost to finance Z ; hence, this could not be globally optimal.

Proof of Proposition 2

To obtain the optimal level of the nudge, we use Z as the only choice variable, for any given level of t . The first order condition for the social optimum is

$$n[-u_C(\eta D_P + q) + u_D \eta D_P + n u_E E'(nD) \eta D_P] = 0. \quad (A4)$$

Substitute in the individual optimum condition (A1) in (A4) and solve for Z :

$$Z = \frac{\frac{q}{\eta D_P} + n MRS_{EC} E'(nD) - t}{\eta} = \frac{\frac{q}{D_Z} + t^{Pigou} - t}{\eta} \quad (A5)$$

To obtain the optimal level of the tax, we use the social first order condition with respect to the tax level:

$$-D_P + MRS_{DC} D_P + n MRS_{EC} E'(nD) D_P = 0, \quad (A6)$$

and substitute in the individual optimal condition (A1) into (A6) and solve for t :

$$t = -\eta Z - n MRS_{EC} E'(nD) = t^{Pigou} - \eta Z. \quad (A7)$$

Proposition 3

The individual optimum condition implies that

$$MRS_{DC} + MRS_{MC} M_D = 1 + t. \quad (A8)$$

The social optimum, where the externality is taken into account, implies that

$$MRS_{DC} + MRS_{MC} M_D + n \frac{u_E E'(nD)}{u_C} = 1. \quad (A9)$$

Combining the private and the social optimum condition we have

$$t = -n \frac{u_E E'(nD)}{u_C} = -n MRS_{EC} E'(nD) = t^{Pigou}. \quad (A10)$$

To obtain the optimal level of the nudge given a Pigovian we substitute the resource constraint into the utility function and look at the first order condition for the social optimum:

$$n[-(D_Z + q) + MRS_{DC} D_Z + n MRS_{EC} E'(nD) D_Z + MRS_{MC} (M_Z + M_D D_Z)] = 0 \quad (A11)$$

Substitute the individual first order condition (A1) and the optimal tax expression (A10) into (A11):

$$-\frac{q}{D_Z} + MRS_{MC} \frac{M_Z}{D_Z} = 0.$$

This implies that:

$$q = MRS_{MC} M_Z. \quad (A12)$$

An interior solution is only possible for $M_Z > 0$, otherwise $Z = 0$.

Proof of Proposition 4

The first order condition for the socially optimal level of the nudge is given by (A11). Substituting the individual first order condition (A1) into (A11), we obtain

$$-q + tD_Z + nMRS_{EC} E'(nD)D_Z + MRS_{MC} M_Z = 0, \quad (A13)$$

implying that the condition for the optimal level of the nudge is, for $Z > 0$,

$$-\frac{q}{D_Z} = t^{Pigou} - t - MRS_{MC} \frac{M_Z}{D_Z}. \quad (A14)$$

The first order condition for an optimal tax, given any level of the nudge, is given by:

$$n \left[-u_C D_{p^D} + u_D D_P + nu_E E'(nD) D_P + u_M M_D D_P \right] = 0. \quad (A15)$$

Substituting the individual first order condition (A15) we obtain:

$$t = -nMRS_{EC} E'(nD) = t^{Pigou}. \quad (A16)$$

Proof of Proposition 5

Combining the private, $MRS_{DC} = 1 + t - b + \eta Z$, and social optimum conditions, $MRS_{DC} + nMRS_{EC} E'(nD) = 1$, we have:

$$1 + t - b + \eta Z = 1 - nMRS_{EC} E'(nD). \quad (A15)$$

It is clear that this condition is satisfied when:

$$t = b - nMRS_{EC} E'(nD) = b + t^{Pigou}, \quad (A16)$$

and $Z = 0$. It would also be satisfied with other combinations where $Z > 0$, but that would be associated with a resource cost to finance Z ; hence, this could not be globally optimal. To find the level of the nudge when the tax is not optimal, substitute the resource constraint into the utility function and derive the first order condition:

$$n \left[-(\eta D_P + q) + MRS_{DC} \eta D_P + nMRS_{EC} E'(nD) \eta D_P \right] = 0. \quad (A17)$$

Substitute the individual first order condition, $MRS_{DC} = 1 - b + t + \eta Z$, and solve for Z :

$$Z = \frac{\frac{q}{D_Z} + b + t^{Pigou} - t}{\eta}. \quad (A18)$$

The first order condition for an optimal tax given any level of the nudge is given by:

$$-1 + MRS_{DC} + nMRS_{EC} E'(nD) = 0. \quad (A19)$$

Substitute the individual first order condition, $MRS_{DC} = 1 - b + t + \eta Z$, and solve for t :

$$t = -nMRS_{EC} E'(nD) + b - \eta Z = t^{Pigou} + b - \eta Z. \quad (A20)$$

Proposition 6

Combining the individual optimum condition, $MRS_{DC} + MRS_{MC} M_D = 1 + t - b$, and the social optimum condition, $MRS_{DC} + MRS_{MC} M_D + nMRS_{MC} E'(nD) = 1$, we have:

$$t = -nMRS_{EC} E'(nD) + b = t^{Pigou} + b. \quad (A21)$$

To find the optimal level of a nudge when the tax is optimal, substitute the resource constraint into the utility function and derive the first order conditions:

$$-(D_Z + q) + MRS_{DC}D_Z + nMRS_{EC}E'(nD)D_Z + MRS_{MC}(M_Z + M_D D_Z) = 0. \quad (A22)$$

Substitute the individual first order condition, $MRS_{DC} + MRS_{MC}M_D = 1 + t - b$, and the optimal tax in (A21) into the first order condition:

$$\begin{aligned} -q + (-b - nMRS_{EC}E'(nD) + b)D_Z + nMRS_{EC}E'(nD)D_Z + MRS_{MC}M_Z \\ = -q + MRS_{MC}M_Z = 0 \end{aligned} \quad (A23)$$

An interior solution only possible for $M_Z > 0$, otherwise $Z = 0$. The social first order condition for the nudge when the tax is not optimal is now:

$$-(D_Z + q) + MRS_{DC}D_Z + nMRS_{EC}E'(nD)D_Z + MRS_{MC}(M_Z + M_D D_Z) = 0. \quad (A24)$$

Substitute the individual first order condition, $MRS_{DC} = 1 - b + t - MRS_{MC}M_D$, and rearrange:

$$\frac{-q}{D_Z} = -nMRS_{EC}E'(nD) - t + b - MRS_{MC} \frac{M_Z}{D_Z} = t^{Pigou} - t + b - MRS_{MC} \frac{M_Z}{D_Z}. \quad (A25)$$

The first order condition for an optimal tax given an existing level of the nudge is:

$$-1 + MRS_{DC} + nMRS_{EC}E'(nD) + MRS_{MC}M_D = 0. \quad (A26)$$

Substitute the individual first order condition $MRS_{DC} = 1 - b + t - MRS_{MC}M_D$ and rearrange to obtain the optimal level of the tax:

$$t = -nMRS_{EC}E'(nD) + b = t^{Pigou} + b. \quad (A27)$$

Proposition 7

The individual optimum condition implies that:

$$MRS_{DC} + \frac{1-\alpha}{1-\beta} CMRS_{MC}M_D = 1 + t. \quad (A28)$$

The social optimum condition implies that:

$$MRS_{DC} + CMRS_{MC}M_D + n CMRS_{EC}E'(nD) = 1. \quad (A29)$$

Combining these two and solve for the optimal tax t we have:

$$t = -nCMRS_{EC}E'(nD) + \frac{\beta-\alpha}{1-\beta} CMRS_{MC}M_D = t^{PigouC} + \frac{\beta-\alpha}{1-\beta} CMRS_{MC}M_D \quad (A30)$$

This can also be written as:

$$\begin{aligned} t &= -n \frac{MRS_{EC}}{1-\alpha} E'(nD) + \frac{\beta-\alpha}{(1-\alpha)(1-\beta)} MRS_{MC}M_D \\ &= \frac{t^{Pigou}}{1-\alpha} + \frac{\beta-\alpha}{(1-\alpha)(1-\beta)} MRS_{MC}M_D \end{aligned} \quad (A31)$$

To find the optimal level of the nudge substitute the resource constraint into the utility function and derive the social first order condition:

$$-(D_Z + q) + MRS_{DC}D_Z + nCMRS_{EC}E'(nD)D_Z + CMRS_{MC}(M_Z + M_D D_Z) = 0 \quad (A32)$$

Substitute the individual first order condition in (A28) and the optimal tax in (A30) and rearrange:

$$-q + CMRS_{MC}M_Z = 0 \quad (A33)$$

Interior solution only possible for $M_Z > 0$, otherwise $Z = 0$.

Proposition 8

Same socially optimal first order condition for the nudge as in (A32). The individual first order condition is now:

$$MRS_{DC} + \frac{1-\alpha}{1-\beta} CMRS_{MC} M_D = 1+t. \quad (A34)$$

Substitute this into social first order condition and rearrange:

$$-q + tD_Z - \frac{\beta-\alpha}{1-\beta} CMRS_{MC} M_D D_Z + nCMRS_{EC} E'(nD) D_Z + CMRS_{MC} M_Z = 0. \quad (A35)$$

Then:

$$\begin{aligned} -\frac{q}{D_Z} &= -t - nCMRS_{EC} E'(nD) + \frac{\beta-\alpha}{1-\beta} CMRS_{MC} M_D - CMRS_{MC} \frac{M_Z}{D_Z} \\ &= t^{Pigou} - t + \frac{\beta-\alpha}{1-\beta} CMRS_{MC} M_D - CMRS_{MC} \frac{M_Z}{D_Z} \end{aligned} \quad (A36)$$

To find the optimal tax for a given level of the nudge, substitute the individual first order condition into social first order condition for a tax and solve for t:

$$t = -nCMRS_{EC} E'(nD) + \frac{\beta-\alpha}{1-\beta} CMRS_{MC} M_D. \quad (A37)$$

Proposition 9

Without loss of generality, we can write the individual budget constraint as:

$$C^i = Y^i - D^i - qZ + \gamma^i. \quad (A38)$$

where $\sum_{i=1}^n \gamma^i = 0$ and where in equilibrium $\gamma^i = -\tau^i - tD^i + qZ$, where τ^i is the individual lump-sum tax. Substitute the resource constraint in the social welfare function (23) using the utility function in (18):

$$\begin{aligned} W &= -QN + \sum_{i=1}^n \gamma^i + \sum_{i=1}^n Y^i - D^i + g^i(D^i, E) \\ &= -QN + \sum_{i=1}^n Y^i - D^i + g^i(D^i, E) \end{aligned} \quad (A39)$$

The first order condition for the socially optimal level of the tax is:

$$\sum_{i=1}^n (-D_P^i + g_{D^i}^i D_P^i) + \sum_{i=1}^n g_E^i E_D \sum_{j=1}^n D_P^j = 0. \quad (A40)$$

This can also be written as:

$$\sum_{i=1}^n (-D_P^i + MRS_{DC}^i D_P^i) + \sum_{i=1}^n MRS_{EC}^i E_D \sum_{j=1}^n D_P^j = 0. \quad (A41)$$

The individual first order condition implies that:

$$MRS_{DC}^i = 1+t - b^i + \eta^i Z, \quad (A42)$$

Substitute this in (A41) and rearrange

$$(t - t^{Pigou}) \sum_{j=1}^n D_P^j - \sum_{j=1}^n b^j D_P^j + Z \sum_{i=1}^n \eta^i D_P^i = 0. \quad (A43)$$

Rewrite this using normalized covariances:

$$(t - t^{Pigou}) n \overline{D_{p^D}} - n \overline{b} \overline{D_{p^D}} (1 + \delta_{b, D_p}) + n Z \overline{\eta} \overline{D_{p^D}} (1 + \delta_{\eta, D_p}) = 0 \quad (A44)$$

Which means that the optimal tax is given by:

$$t = t^{Pigou} + \overline{b} (1 + \delta_{b, D_p}) - Z \overline{\eta} (1 + \delta_{\eta, D_p}) \quad (A45)$$

Next derive the first order condition for the socially optimal level of the nudge:

$$-Q + \sum_{i=1}^n \eta^i (-D_P^i + g_{D^i}^i D_P^i) + \sum_{i=1}^n g_{E_D}^i E_D \sum_{j=1}^n \eta^j D_P^j = 0 \quad (\text{A46})$$

Or

$$-Q + \sum_{i=1}^n \eta^i (-D_P^i + MRS_{DC}^i D_P^i) + \sum_{i=1}^n MRS_{EC}^i E_D \sum_{j=1}^n \eta^j D_P^j = 0 \quad (\text{A47})$$

Substitute the individual optimum condition in (A42) and rearrange

$$\begin{aligned} -Q + t \sum_{i=1}^n \eta^i D_P^i - \sum_{i=1}^n b^i \eta^i D_P^i + \sum_{i=1}^n \eta^i \eta^i Z D_P^i \\ + \sum_{i=1}^n MRS_{EC}^i E_D \sum_{j=1}^n \eta^j D_P^j = 0 \end{aligned} \quad (\text{A48})$$

Use the short notation $D_N^k \equiv \eta^k D_P^k$ and normalized covariances and solve for Z:

$$Z = \frac{\frac{Q}{nD_Z} - \left(t + \sum_{i=1}^n MRS_{EC}^i E_D - \bar{b}(1 + \delta_{b,D_Z}) \right)}{\bar{\eta}(1 + \delta_{\eta,D_Z})} = \quad (\text{A49})$$

$$\frac{\frac{q}{D_Z} + t^{Pigou} - t + \bar{b}(1 + \delta_{b,D_Z})}{\bar{\eta}(1 + \delta_{\eta,D_Z})}$$

This is the optimal level of the nudge for any given level of the tax. Substitute the optimal tax in (A49) and solve for Z:

$$Z = \frac{\frac{q}{D_Z} + \bar{b}(\delta_{b,D_Z} - \delta_{b,D_P})}{\eta(\delta_{\eta,D_Z} - \delta_{\eta,D_P})} \quad (\text{A51})$$

Proposition 10

Follows from 9

Proposition 11 and 12

The social optimum condition for the nudge implies that:

$$\begin{aligned} -Q + \sum_{i=1}^n (-D_Z^i + MRS_{DC}^i D_Z^i) + \sum_{i=1}^n MRS_{EC}^i E_D \sum_{j=1}^n D_Z^j \\ + \sum_{i=1}^n MRS_{MC}^i (M_Z^i + M_D^i D_Z^i) = 0 \end{aligned} \quad (\text{A52})$$

The individual perceived optimum condition implies that

$$MRS_{DC}^i = 1 + b^i + t - MRS_{MC}^i M_D^i \quad (\text{A53})$$

Substitute (A52) in (A53):

$$-Q + \left(t + \sum_{i=1}^n MRS_{EC}^i E_D \right) n \bar{D}_Z - \sum_{i=1}^n b^i D_Z^i + \sum_{i=1}^n MRS_{MC}^i M_Z^i = 0. \quad (\text{A54})$$

Use short notations for the normalized covariances

$$\delta_{MRS_{MC}, M_Z} \equiv \text{cov} \left(\frac{MRS_{MC}}{MRS_{MC}}, \frac{M_Z}{M_Z} \right) \text{ and } \delta_{b, D_Z} \equiv \text{cov} \left(\frac{b}{b}, \frac{D_Z}{D_Z} \right) \quad (\text{A55})$$

Implying

$$\begin{aligned} -Q + \left(t + \sum_{i=1}^n MRS_{EC}^i E_D \right) n \bar{D}_Z - n \bar{b} \bar{D}_Z (1 + \delta_{b, D_Z}) + \\ n \overline{MRS_{MC}} \overline{M_Z} (1 + \delta_{MRS_{MC}, M_Z}) = 0 \end{aligned} \quad (\text{A56})$$

So the condition for the optimal level of the nudge for a non-optimal level of the tax is:

$$-\frac{q}{D_Z} = t^{Pigou} - t + \bar{b}(1 + \delta_{b,D_Z}) + \frac{\overline{MRS_{MC}M_Z}}{D_Z} \quad (A57)$$

Next, take social first order conditions with respect to the tax:

$$\begin{aligned} \sum_{i=1}^n (-D_P^i + MRS_{DC}^i D_P^i) + \sum_{i=1}^n MRS_{EC}^i E_D \sum_{j=1}^n D_P^j \\ + \sum_{i=1}^n MRS_{MC}^i M_D^i D_P^i = 0 \end{aligned} \quad (A58)$$

Substitute the individual perceived optimum condition, $MRS_{DC}^i = 1 - b^i + t - MRS_{MC}^i M_D^i$, and use short notation for normalized covariance and solve for the optimal tax:

$$t = t^{Pigou} + \bar{b}(1 + \delta_{b,D_P}) \quad (A59)$$

Substitute the optimal tax in (A59) into the expression for optimal level of nudge in (A57) and rearrange

$$-\frac{q}{D_Z} = (\delta_{b,D_Z} - \delta_{b,D_P})\bar{b} + \frac{\overline{MRS_{MC}M_Z}}{D_Z} \quad (A60)$$

Propositions 13 and 14

Write the SWF as:

$$\begin{aligned} W = -QN + \sum_{i=1}^n Y^i - D^i \\ + g^i(Y^i - \bar{Y} - D^i + \bar{D} + \alpha^i, D^i, D^i - \bar{D}, E, M^i(D^i, D^i - \bar{D}, Z)) \end{aligned} \quad (A61)$$

Take first order condition with respect to the nudge:

$$\begin{aligned} -Q - n\bar{D}_Z + \frac{1}{1-\alpha} \sum_{i=1}^n MRS_{DC}^i D_Z^i - \frac{\alpha}{1-\alpha} \bar{D}_Z \sum_{i=1}^n MRS_{DC}^i + n\bar{D}_Z \sum_{i=1}^n MRS_{EC}^i E_D \\ + \sum_{i=1}^n MRS_{MC}^i (M_Z^i + M_D^i D_Z^i + M_{D^i-\bar{D}}^i D_Z^i) - \bar{D}_Z \sum_{i=1}^n MRS_{MC}^i M_{D^i-\bar{D}}^i = 0 \end{aligned} \quad (A60)$$

The private first order condition gives us:

$$MRS_{DC}^i = 1 + b^i + t - \frac{g_M^i(M_D^i + M_{D^i-\bar{D}}^i)}{1 + g_{C^i-\bar{C}}^i} \quad (A60)$$

Substitute:

$$\begin{aligned} -Q + tn\bar{D}_Z + \frac{1}{1-\alpha} \sum_{i=1}^n b^i D_Z^i - \frac{\alpha}{1-\alpha} \bar{D}_Z \sum_{i=1}^n B^i + n\bar{D}_Z \sum_{i=1}^n CMRS_{EC}^i E_D \\ + \sum_{i=1}^n CMRS_{MC}^i M_Z^i - \bar{D}_Z \sum_{i=1}^n \frac{\beta^i - \alpha}{1 - \beta^i} CMRS_{MC}^i M_{D^i}^i = 0 \end{aligned} \quad (A60)$$

Use the short notation $\delta_{MRS_{MC}, M_Z} \equiv \text{cov}\left(\frac{MRS_{MC}}{MRS_{MC}}, \frac{M_Z}{M_Z}\right)$ and $\delta_{b, D_Z} \equiv \text{cov}\left(\frac{b}{b}, \frac{D_Z}{D_Z}\right)$:

$$\begin{aligned} -Q + tn\bar{D}_Z + n\bar{D}_Z \bar{b} \left(1 + \frac{\delta_{D_N, b}}{1-\alpha}\right) + n\bar{D}_Z \frac{t^{Pigou}}{1-\alpha} + \sum_{i=1}^n CMRS_{MC}^i M_Z^i \\ - n\bar{D}_Z \left(\frac{\beta - \alpha}{1 - \beta}\right) \overline{CMRS_{MC}M_D} \left(1 + \delta_{\frac{\beta - \alpha}{1 - \beta}, CMRS_{MC}M_D}\right) = 0 \end{aligned} \quad (A60)$$

Rearranging we can express the condition for the optimal level of the nudge as:

$$-\frac{q}{D_Z} = t^{PigouC} - t - \bar{b} \left(1 + \frac{\delta_{D_N, b}}{1 - \alpha} \right) - \frac{\overline{CMRS_{MC} M_Z}}{D_Z} + \left(\frac{\beta - \alpha}{1 - \beta} \right) \overline{CMRS_{MC} M_D} \left(1 + \delta_{\frac{\beta - \alpha}{1 - \beta}, CMRS_{MC} M_D} \right) \quad (A60)$$

Consider next the optimal tax, and take the first order condition w.r.t. the tax:

$$-n\bar{D}_P + \frac{1}{1 - \alpha} \sum_{i=1}^n MRS_{DC}^i D_P^i - \frac{\alpha}{1 - \alpha} \bar{D}_{p^D} \sum_{i=1}^n MRS_{DC}^i + n\bar{D}_Z \sum_{i=1}^n MRS_{EC}^i E_D + \sum_{i=1}^n MRS_{MC}^i (M_D^i D_P^i + M_{D^i - \bar{D}}^i D_P^i) - \bar{D}_P \sum_{i=1}^n MRS_{MC}^i M_{D^i - \bar{D}}^i = 0 \quad (A60)$$

Write the individual first order condition as:

$$MRS_{DC}^i = 1 + b^i + t - \frac{g_M^i (M_D^i + M_{D^i - \bar{D}}^i)}{1 + g_{C^i - \bar{C}}^i} = 1 + b^i + t - (1 - \alpha) g_M^i (M_D^i + M_{D^i - \bar{D}}^i) \quad (A60)$$

Substitute and rearrange

$$n\bar{D}_P t - n\bar{D}_P \frac{t^{Pigou}}{1 - \alpha} + n\bar{D}_P \bar{b} \left(1 + \frac{\delta_{D_P, b}}{1 - \alpha} \right) + \bar{D}_P \sum_{i=1}^n \frac{\alpha - \beta^i}{1 - \beta^i} CMRS_{MC}^i M_{D^i}^i = 0 \quad (A60)$$

Solving for the tax we have

$$t = t^{PigouC} - \bar{b} \left(1 + \frac{\delta_{D_P, b}}{1 - \alpha} \right) - \left(\frac{\beta - \alpha}{1 - \beta} \right) \overline{CMRS_{MC} M_D} \left(1 + \delta_{\frac{\beta - \alpha}{1 - \beta}, CMRS_{MC} M_D} \right) \quad (A60)$$

Substitute the tax expression in the condition for the optimal level of the nudge and rearrange:

$$-\frac{q}{D_Z} = \frac{\bar{b}}{1 - \alpha} (\delta_{D_P, b} - \delta_{D_Z, b}) - \frac{\overline{CMRS_{MC} M_Z}}{D_Z} \quad (A60)$$