GREEN LIFESTYLES AND SOCIAL TIPPING POINTS

Documents de travail GREDEG
GREDEG Working Papers Series

Michael Finus
Paolo Zeppini

GREDEG WP No. 2026-08

https://ideas.repec.org/s/gre/wpaper.html

Green Lifestyles and Social Tipping Points

Michael Finus[1,3] and Paolo Zeppini[2] *

March 3, 2026

[1] School of Business, Economics and Social Sciences, University of Graz, Austria
[2] Université Côte d'Azur, CNRS, GREDEG, France
[3] Department of Economics, University of Bath, United Kingdom

Abstract

We introduce the concept of *green lifestyles* in an economic discrete choice model of consumption behaviour. Agents behave in either a 'selfish' or 'pro-social' way by choosing different degrees of internalisation of environmental damage from the consumption of an environmentally harmful good. Pro-social behaviour means lower consumption, and is rewarded with warm-glow. Moreover, the agents' decision is influenced by social norms, which endogenously depend on aggregate choices. The model is developed in a dynamic framework, allowing agents to switch behaviour. Our results show that conventional measures limiting consumption at an individual level may increase consumption at the aggregate level. We characterise social tipping points for sustainability transitions in terms of equilibria bifurcations and hysteresis of population dynamics. The model is extended in different directions, with different types of social influence and with a state dependent warm-glow. This more complicated decision environment gives alternative regimes with either dampening or self-reinforcing feedback in decisions. Three scenarios are identified: for strong social norms positive feedback leads to multiple equilibria. For moderate social norms there is a unique equilibrium. For weak social norms, we obtain periodic dynamics of behaviours. In particular, more informed choices and lower variability across agents are 'destabilising', leading to periodic dynamics or multiple equilibria.

JEL classification: C62, D62, Q56.

Keywords: discrete choice, social interactions, sustainable consumption, transitions, warm-glow.

---

[*] Emails: paolo.zeppini@univ-cotedazur.fr (corresponding author), michael.finus@uni-graz.at. Support from the EU FP7 Project GLAMURS is gratefully acknowledged. This article benefited from discussions with Lucy O'Shea, Tobias Wenzel and Maik Schneider. Earlier versions of the paper were presented at seminars and conferences in Bath, Bologna, Manchester, Amsterdam, Zurich, Nice, Osaka and Barcelona. All flaws and errors are ours.

1 Introduction

Environmental pollution and climate change are presently among the most pressing challenges for human mankind. In order to avoid environmental disaster, the United Nations Framework Convention on Climate Change (UNFCCC) has set the target of keeping the global mean temperature increase below 2^oC. The real challenge for policy action is how to achieve this target. The European Environmental Agency has committed to a reduction of Greenhouse Gases Emissions (GHG) of 40% by 2030 and 95% by 2050, compared with 1990 levels.

Both the supply side and the demand side of the economy need to engage in GHG emission reduction. Regarding the supply side, the main way to reduce emission and maintain economic growth is through increased efficiency in terms of output or services produced per unit of GHGs emitted, the most clear example being energy efficiency. However, a good deal of the responsibility for environmental impact lies with private consumers. It has been estimated that household emissions account for a share of 74% of total emissions in UK (Baiocchi et al., 2010).

The demand side of the problem is possibly more subtle and challenging in the quest for a sustainable economy. Beside the adoption of more efficient products and services, a reduction of emissions can also be obtained from a substitution effect towards more 'green' products, or by a direct reduction of consumption of a given harmful good. The individual consumption decision is the outcome not only of economic considerations, but also a domain where psychological aspects and social norms play an important role.

This paper proposes a discrete choice model of consumption decision where individual agents are confronted with a choice between a 'green' and a 'brown' lifestyle, which are defined respectively by a pro-social - or 'environmental' consumption behaviour, and, on the other hand, by a selfish - or 'non-environmental' - behaviour. The crucial role of people's lifestyles in a green transition is recognised within policy making (see for instance the policy position of Friends of the Earth)[1]. By using the notion of social tipping points, we show how such a transition can result from the critical value of some key factors in collective decision making.

In the model we give attention to a number of different motives in the consumption decision, such as the material reward stemming from the satisfaction of physical needs, the - negative - environmental externality from consumption, a psychological well-being from adhering to a personal norm, referred in the literature as 'warm-glow' and, finally, the effect of social descriptive norms that account for the influence of observing what the others do in terms of consumption behaviours.

---

[1] Friends of the Earth is an international network of environmental organisations, `https://policy.friendsoftheearth.uk/opinion/government-has-key-role-shifting-us-green-lifestyles`

The effect of social norms in influencing environmentally relevant behaviour has long been recognised in social psychology (Bell et al., 1990). There are different types of social norms (Bicchieri, 2006). Among them, the case of 'descriptive' social norms is the one the more closely expresses the idea of lifestyles, which is central to our paper. Descriptive norms are normative behaviour that involve what people 'do' rather than what they 'ought' to do. The latter is the realm of 'prescriptive' norms, which are conveyed by rules that are explicitly brought out in public institutions, such as the laws in a legal system and signs in public places. Descriptive norms are instead perceived by an individual when they observe what others do. Descriptive norms are different from 'injunctive' norms, which derive from what people are 'expected' to do, and always carry a form of social reward, negative or positive. In our model we focus on descriptive norms, following the literature on *social interactions* (Manski, 2000; Brock and Durlauf, 2001) that describes them as positive spillovers of behaviours. In psychology this phenomenon has been addressed in studies of social influence and conformity (Cialdini and Goldstein, 2004).

Beside social norms, an individual behaviour is influenced by *personal norms*. There is a link between social and personal norms. An individual "feels good" about herself when her action conforms to some moral ideal which links to environmental impact (Brekke et al., 2003). Personal norms are then social norms that have already been internalised within the individual. In our context it can be interpreted as to what is referred to as *warm-glow* in the economics literature (Andreoni, 1990, 1995; Van der Linden, 2018) and as intrinsic motivation or *psychological well-being* in the psychological literature (Ryan and Deci, 2000). Andreoni observed that government funding of charitable donations only imperfectly crowds out private donations. The reason being that individuals derive utility from the donation itself. Warm-glow is then a form of 'impure' altruism, as it does not stem from the utility of others. Extending Andreoni's idea to the environmental sphere we model warm-glow as the psychological well-being derived from a 'green' choice of sustainable consumption - say for instance organic food. Warm-glow can then be regarded as an 'extra' motivation to engage in environmentally friendly behaviour when one is not seen (reputation), does not expect anything back (reciprocity), and is not caring for others (pure altruism).

Environmentally friendly behaviour is what makes a 'green' lifestyle in our model. Lifestyles concern behaviour that is embedded in a social context, and shared by a population of individuals. An ideal theoretical framework for such a model is social interactions and discrete choice, as in Brock and Durlauf (2001). We propose this framework for the study of environmentally-relevant behaviour and its psychological and sociological drivers, and extend this model with a micro-foundation of discrete choice utility from consumption optimisation, within a two-stages decision process. In particular we consider a dynamic setting with a market equilibrium of supply and demand of the environmentally relevant good, and a price adjustment mechanism, similar to a tâtonnement process, which in the discrete choice framework was introduced by Brock and Hommes (1997). In this way we model the interplay of individual consumption decisions and social interactions, describing the effect of more or less sustainable lifestyles.

Behaviours in our model are instantiated as consumption decisions about an environmentally harmful good. The drivers of decisions are, first, material reward, second, environmental damage, third, warm-glow, and finally social interactions. The model describes a discrete choice between a green lifestyle and a brown lifestyle. The brown lifestyle is defined as a 'selfish' choice, a consumption level with a relatively low perceived marginal environmental damage. The green lifestyle is a 'social' one instead, with a relatively high perceived marginal environmental damage. In particular, the two marginal damage levels can be set to zero and to the full internalisation of overall consumption, respectively.

Lifestyle choices are repeated in time, with two stages of the decision process at each time interval. In the first stage, agents decide which lifestyle to adopt among the two possibilities. In the second stage, conditional on the lifestyle choice, agents optimise their consumption level. The two stages are similar in a way to the setting of Brekke et al. (2003), where in place of an 'ideal' choice for the social norm we have two opposed choices, defining competing lifestyles. In our model we adopt a population approach where individual decisions are aggregated in an evolutionary fashion through the fraction of individuals adopting the 'green' behaviour (or sustainable lifestyle), as opposed to the fraction of individuals adopting the 'brown' behaviour (unsustainable lifestyle). The fraction of behaviours is a state variable that feeds back into the stage one consumption decision through a market equilibrium where total demand across the population is paired to the supply of such a harmful good.

We introduce and mathematically describe the concept of social tipping points, as discontinuities and sudden changes in aggregate behaviours over lifestyle choice. There are two types of social tipping points: the first type works as a critical mass effect of one lifestyle adoption, and is identified by the possible unstable equilibrium separating two stable equilibria. The second type of social tipping point is the transition between two different equilibrium scenarios, induced by the change in a parameter crossing a critical value (bifurcation). Moreover, our results uncover surprising outcomes such as the effect of reducing marginal damage levels: although consumption decreases at the individual level, it increases at the aggregate level.

The model is extended in two directions, first with different types of social influence and then with a warm-glow dependent on average choices. In summary, three scenarios are identified: for strong social interactions positive feedback leads to multiple equilibria. For moderate social interactions there is a unique equilibrium. For weak social interactions, we obtain periodic dynamics of behaviours. In particular, more informed choices and lower variability across agents are 'destabilising', leading to periodic dynamics or multiple equilibria.

The article is organised as follows. Section 2 explains the general model of sustainable lifestyles. Section 3 https://www.overleaf.com/project/5e5f88979a303f0001761388presents a basic version of the model and focuses on social interactions. Sections 4 and 5 extend the model with crowding-out in social interactions and with endogenous warm-glow, respectively. Section 7 concludes.

2 A Discrete Choice Model of Consumption Lifestyles

In this section, we set out the general features of our model and report about some general results, which we use in subsequent sections.

We consider a population of individuals who take repeatedly two decisions at each point in time t:

(1) *choice of a lifestyle* and

(2) *choice of consumption level.*

In each stage, the game is solved by backward induction. In the first stage, individuals face a discrete choice between two lifestyles, 'brown' and 'green', which are associated with two alternative behaviours:

- a *brown* lifestyle, which may be viewed as 'selfish' behaviour, and
- a *green* lifestyle, which may be viewed as 'pro-social' behaviour.

Given the choice in the first stage, individuals face a continuum of consumption levels of an environmentally harmful good in the second stage. We assume that individuals who choose the green lifestyle (henceforth referred to as 'greens') internalise a larger fraction of marginal damages caused by consumption than individuals who choose the brown lifestyle (henceforth referred to as 'browns'). That is, greens consume less than browns.[2]. We only consider pure strategies in each of the two stages.

We call the game at time t the *constituent game.* In contrast to ordinary repeated games, in discrete choice models, the game is not the same at each point in time because

---

[2] Different behaviours could also be operationalised by greens (browns) consuming an environmentally friendly (harmful) version of a good.

the state variable will change. Individuals may revise their decisions; there is the possibility that decisions gradually reach a steady state at some point in time, in which case we talk about an equilibrium. However, there may also be cyclical patterns without a steady state equilibrium.

We assume a (large) population of n individuals, $i = 1, \ldots, n$, where individual i's payoff is given by

$$V_i = U_i - pq_i - D_i + G_i + I_i. \tag{1}$$

U_i is the utility derived from consumption of an environmentally harmful good with quantity q_i with the following properties:

$$U_i \equiv U(q_i), \quad U'(q_i) > 0, \quad U''(q_i) < 0. \tag{2}$$

That is, consumption generates a material reward. Utility is strictly increasing and concave in consumption. The expenditure on consumption is given by pq_i where p is the price. Other goods are assumed to be not environmentally harmful. They constitute the *numeraire* good with normalised price 1 and expenditure $y - pq_i$ with y being income. Given that we assume a quasi-linear utility function, we can ignore the numeraire good from now onwards.

D_i is the damage from which consumers suffer, which we assume is a function of total consumption $Q = \sum_{k=1}^{n} q_k$. That is, we assume a constant emission output ratio which, without loss of generality, we normalise to 1. Emissions released in consumption constitute a pure public bad. Damages increase in total emissions/output at an increasing or constant rate:

$$D_i = D(Q), \quad D'(Q) > 0, \quad D''(Q) \geq 0. \tag{3}$$

The remaining two terms in the payoff function relate to norms. G_i is the utility which an individual derives from pursuing a green lifestyle, which is associated with a *warm-glow* or *psychological well-being.* In contrast, I_i is the utility which an individual derives from adhering to a *social norm.* Social utility depends on the lifestyle that an individual adopts and the distribution of lifestyles in the population. We will consider different assumptions about what drives psychological well-being and social norms. They will be explained in more detail in subsequent sections.

As it will become apparent successively, all components of the payoff function depend not only on own decisions, but also on the decision of all individuals in the population. That is, there are spillovers or externalities across individuals through various channels.

In the second stage, there is a fraction of individuals x who have adopted the green lifestyle, consuming quantity q_g, and a fraction $1-x$ who have adopted the brown lifestyle, consuming quantity q_b where we drop the time index t for notational simplicity. The maximisation of payoff (1) delivers equilibrium consumption levels for greens and browns.

It turns out that $q_g^* < q_b^*$ in all our model versions, which we consider subsequently, where we also explain in more detail the micro-foundation of different consumption choices.

Total demand for a given x and resulting equilibrium price p^* is $Q^*(p^*(x)) = \sum_{k=1}^{n} q_k^*(p^*(x)) = n[xq_g^*(p^*(x)) + (1-x)q_b^*(p^*(x))]$. Individual (as well as total) demand is assumed to be downward sloping, i.e., $\frac{\partial q_i^*}{\partial p} < 0$ ($\frac{\partial Q^*}{\partial p} < 0$). Moreover, we assume an upward sloping supply curve $S(p)$, i.e., $\frac{\partial S(p)}{\partial p} > 0$. Hence, the market equilibrium is given by $p^*(x) = S^{-1}[Q(p^*(x))]$. This has the following implications.

Proposition 2.1. *Total equilibrium consumption and the equilibrium price decrease with the fraction of individuals who choose a green lifestyle x, i.e., $\frac{\partial Q^*}{\partial x} < 0$ and $\frac{\partial p^*}{\partial x} < 0$.*

Proof. We have $\frac{\partial Q^*}{\partial x} = n(q_g^* - q_b^* + x\frac{\partial q_g^*}{\partial x} + (1-x)\frac{\partial q_b^*}{\partial x})$. We note that $\frac{\partial q_g^*}{\partial x} = \frac{\partial q_g^*}{\partial p^*}\frac{\partial p^*}{\partial x}$. With $S(p)$ being invertible, $p^*(x) = S^{-1}[Q(p^*(x))]$, we have $\frac{\partial p^*}{\partial x} = \frac{\frac{\partial Q^*}{\partial x}}{S'[S^{-1}(Q^*)]}$. Substituting into $\frac{\partial Q^*}{\partial x}$, we obtain

$$\frac{\partial Q^*}{\partial x}\left(1 - \frac{x\frac{\partial q_g^*}{\partial p^*} + (1-x)\frac{\partial q_b^*}{\partial p^*}}{S'(S^{-1}(Q^*))}\right) = n(q_g^* - q_b^*) < 0,$$

because $q_g^* < q_b^*$. Since $S'[S^{-1}(Q)] > 0$, we also have $\frac{\partial p^*}{\partial x} < 0$. □

Therefore, changes in lifestyle will affect the total equilibrium quantity consumed and the equilibrium price.

The decision in the first stage of the constituent game at time t of individual $i = 1$ is given by:

$$\omega_i = \begin{cases} 1 & \textit{green} \text{ lifestyle} \\ 0 & \textit{brown} \text{ lifestyle.} \end{cases}$$

The individual payoff depends on the choice ω_i, individual characteristics which are related to a deterministic vector $\mathbf{Z}_i$, as well as random shocks $\epsilon_i(\omega_i)$ (Brock and Durlauf, 2001):

$$\tilde{V}_i(\omega_i, \mathbf{Z}_i, \epsilon_i(\omega_i)) = V_i(\omega_i, \mathbf{Z}_i) + \epsilon_i(\omega_i). \tag{4}$$

The decision in the first stage of individual i is given by:

$$\max_{\omega_i \in \{0,1\}} \tilde{V}_i(\omega_i, \mathbf{Z}_i, \epsilon_i(\omega_i)). \tag{5}$$

A common assumption in discrete choice theory is that noise terms $\epsilon_i(\omega_i)$ are independently and *extreme valued* distributed across individuals. Accordingly, the differences $\epsilon_i(1) - \epsilon_i(0)$ are *logistically* distributed (McFadden, 1981; Brock and Hommes, 1997):

$$\mathit{Prob}\,\{\epsilon_i(1) - \epsilon_i(0) \le z\} = \frac{1}{1 + e^{-\beta z}}, \tag{6}$$

with $\beta \in \mathbb{R}^+$ the *intensity of choice.* The parameter β has several interpretations. In the traditional discrete choice theory, it describes the variability in consumers' tastes, the so-called preference 'shocks' (McFadden, 1981). The larger (smaller) the value of β, the smaller (larger) the variability of tastes. That is, across all agents, but even within a social group, we can capture different preferences through the parameter β. In the discrete choice literature analysing social interactions (Brock and Durlauf, 2001) and heterogeneous expectations (Brock and Hommes, 1997), the parameter β is viewed as a measure of the degree of rationality, i.e., the ability of individuals to form correct beliefs about a state variable and make rational choices. The larger the value of β, the larger the degree of rationality (i.e., the lower the probability of an error). In our context, both interpretations are interesting and meaningful.

The probability $Prob\{\omega_i = 1\}$ of making a 'green' choice and the probability $1 - Prob\{\omega_i = 1\}$ of making a 'brown' choice are also logistically distributed. In the context of our model, with a large population, this probability is equivalent to the fraction of individuals x choosing a green lifestyle and is a function of the difference in deterministic payoffs between a *brown* and a *green* lifestyle $\Delta V(x) = V_b(x) - V_g(x)$. Accordingly, the probability distribution of the fraction x is given by the following logistic map f.

$$x(\hat{x}) = f(\hat{x}) = \frac{e^{\beta V_g(\hat{x})}}{e^{\beta V_g(\hat{x})} + e^{\beta V_b(\hat{x})}} = \frac{1}{1 + e^{\beta \Delta V(\hat{x})}}. \tag{7}$$

The equilibrium value x is a *fixed point*, $x(\hat{x}) = \hat{x}$, implying *consistency of beliefs.* Considering $x = \hat{x}$ only amounts to a steady-state analysis. If we do not make this assumption a priori, then we obtain a dynamic system with possible adjustment over time, like a tâtonnement process (Brock and Hommes, 1997): individuals may switch behaviour over time. A frequently made assumption for the revision protocol of repeated discrete choice is $\hat{x} = x_{t-1}$ (Hommes, 2013; Zeppini, 2015). Hence, at a given time t, an individual evaluates her payoff based on the distribution of choices realised one period before. This allows us to study how the fraction of 'greens' x changes over time, its direction, on which parameters and initial conditions this depends and whether the dynamic system converges over time. In the dynamic context, both stable and unstable equilibria are fixed points. Stable equilibria are those which cannot be upset by small perturbations.

Figure 1 illustrates some possible shapes of the mapping $f(x)$. The examples assume that $V(x)$ is linear in x, which is a condition satisfied in all versions of our model. Hence, also the difference in payoffs between a brown and green lifestyle $\Delta V(x) = V_b(x) - V_g(x)$ is linear in x. This has the following implications.

First, $f(x)$ can be upward or downward sloping. If it is upward sloping, i.e., $f'(x) > 0$ (examples (a), (b), (c) and (d) in Figure 1), the larger the fraction x at time $t-1$, x_{t-1}, the larger will be the fraction x at time t, x_t. That is, lifestyle choices are reinforcing, which is called *positive feedback.* The reverse is true if $f(x)$ is downward sloping, i.e.,

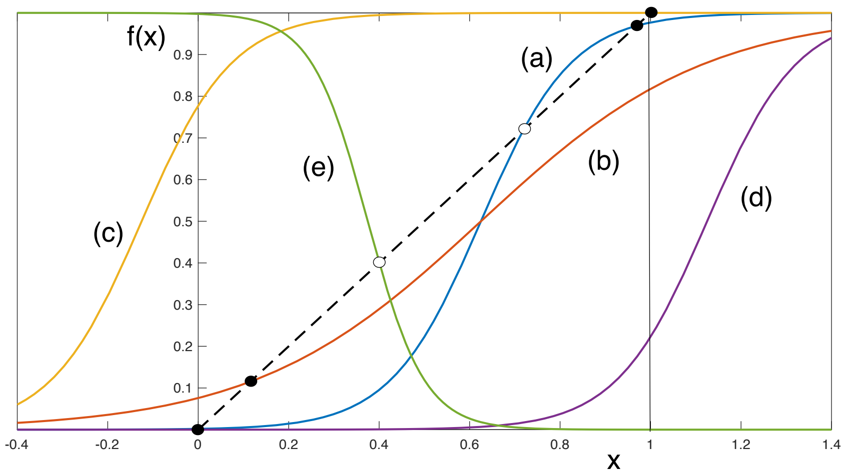

Figure 1: Examples of the function $f(x)$ in equation (7). The dots on the 45^o-line indicate equilibria, with a filled dot for stable and a circled dot for unstable equilibria. a) Increasing distribution $f(x)$ with two stable equilibria and one unstable equilibrium. b) Increasing distribution $f(x)$ with $f' < 1$ everywhere and one stable equilibrium. c) and d) Increasing distributions $f(x)$ with a flex point $x^F \notin [0,1]$ and one stable equilibrium. e) Decreasing distribution $f(x)$ with one equilibrium, which is not stable because $f'(x^*) < -1$ at the equilibrium x^*.

$f'(x) < 0$ (example (e) in Figure 1), to which we refer as *negative feedback*. We note that

$$f'(x) = -\frac{\beta \Delta V'(x) e^{\beta \Delta V(x)}}{\left(1 + e^{\beta \Delta V(x)}\right)^2}. \tag{8}$$

The sign of the first derivative $f'(x)$ depends on the sign of $-\beta \Delta V'(x)$ as $e^{\beta \Delta V(x)} > 0$ where $\Delta V'(x)$ is the first derivative of $\Delta V(x)$. We note that $\Delta V'(x)$ is a constant different from zero, due to our assumption that $\Delta V(x)$ is linear in x. Hence, depending on the model specification, we have either a positive feedback if $\Delta V'(x) < 0$ or a negative feedback if $\Delta V'(x) > 0$.

Second, by the definition of a fixed point, all values of x for which $f(x)$ intersects with the 45^o-line are equilibria, i.e., $x^* = f(x^*)$. Because $f(x)$ is a mapping into $[0,1]$, it must cross the 45^o-line at least once. Hence, an equilibrium always exists.

Third, stable equilibria are those values of x^* to which the state variable x converges in the long-run. A fixed point $x^* = f(x^*)$ is *stable* if small perturbations result in the state variable x returning to the starting value. In dynamic systems that are one-dimensional, such as ours, a fixed point is stable whenever the first derivative is smaller than one, i.e., $|f'(x^*)| < 1$.

Fourth, if $f(x)$ is downward sloping, it crosses the 45^o-line only once. Hence, there can be at most one equilibrium. If $|f'(x^*)| \leq 1$, x^* is the unique stable equilibrium. However, if the reverse is true, the equilibrium is unstable. As we discuss in detail in Section 4, x_t enters a periodic orbit whenever $|f'(x^*)| > 1$, which implies that x constantly jumps between two values in the long-run.

Fifth, if $f(x)$ is upward sloping, we can conclude the following. Because of its S-shape, either $f(x)$ has one fixed point (examples (b), (c) and (d) in Figure 1) or three fixed points (example (a) in Figure 1). If the fixed point is unique, we have a unique stable equilibrium. If there are three fixed points $x_1^* < x_2^* < x_3^*$, then $x_1^* < 1/2$ and $x_3^* > 1/2$ are stable equilibria while x_2^* is an unstable equilibrium.

Sixth, we call a stable equilibrium $x^* < 1/2$ a 'brown equilibrium' and denote it by x_b^*, and a stable equilibrium $x^* > 1/2$ a 'green equilibrium' and denote it by x_g^*.

Seventh, we noted above that a fixed point is stable whenever the first derivative is smaller than one, i.e. $|f'(x^*)| < 1$. Consequently, a sufficient condition for a unique stable equilibrium is that $|f'(x)| < 1$ in the entire domain of x. In the case of a downward sloping function f, $|f'(x)| \leq 1$ is a sufficient condition for unique stable equilibrium.

Eighth, the steepest point of $f(x)$, i.e., the largest value of $|f'(x)|$, is where $f''(x) = 0$. This is the *flex point* x^F of the logistic function $f(x)$. The flex point turns out to be the *indifference point*, i.e., the value of x for which an individual is indifferent between the green and the brown lifestyle, i.e., $\Delta V(x^F) = 0$. This is evident by considering the second derivative of $f(x)$, where we make use of the fact that $\Delta V''(x) = 0$ because $\Delta V(x)$ is linear in x in our model:

$$f''(x) = \frac{\beta^2 e^{\beta \Delta V(x)} \left(\Delta V'(x)\right)^2 \left(e^{\beta \Delta V(x)} - 1\right)}{\left(1 + e^{\beta \Delta V(x)}\right)^3}. \tag{9}$$

As long as $\beta > 0$, $f''(x) = 0$ if and only if $\Delta V(x) = 0$, which is the case for $x = x^F$. Whenever $\Delta V(x) > 0$, f is convex, as $f''(x) > 0$, while for $\Delta V(x) < 0$, $f(x)$ is concave. Moreover, an increasing function $f(x)$ is convex for $x < x^F$ and concave for $x > x^F$. The opposite is true for a decreasing function $f(x)$. We can say that $f(x)$ is S-shaped.

There are three cases: i) $\Delta V(x) > 0$ for all $x \in [0,1]$ which de facto means $x^F > 1$ (example (d) in Figure 1), ii) $\Delta V(x) < 0$ for all $x \in [0,1]$ which de facto means $x^F < 0$ (example (c) in Figure 1) and iii) $\Delta V(x) = 0$ for some $x \in [0,1]$ and thus $x^F \in [0,1]$ (examples (a),(b) and (e) in Figure 1). As the domain of x is $[0,1]$, an indifference point does not exist in cases (i) and (ii). However, cases (i) and (ii) will turn out to be helpful in locating equilibria for the specific model versions which consider in subsequent sections.

As $f(x)$ is S-shaped, a sufficient condition for a unique stable equilibrium is derived from solving $f'(x^F) \leq 1$, which, using (8), gives

$$\left|f'(x^F)\right| \leq 1 \quad \Rightarrow \quad |\Delta V'(x)| \leq \frac{4}{\beta}. \tag{10}$$

In the context of an upward sloping function $f(x)$, there is a second sufficient condition for a unique stable equilibrium. If the flex point x^F is outside the domain $[0,1]$ (as in examples (c) and (d) in Figure 1), then $f(x)$ crosses the 45^o-line only once with a slope less than 1. Hence, we have a unique stable equilibrium.

All observations discussed above are summarised in Proposition 2.2.

Proposition 2.2. *Consider the revision protocol* $f(x)$ *in eq. (7) with* $\hat{x} = x_{t-1}$ *and let* $\Delta V(x) = V_b(x) - V_g(x)$ *be linear in* x*.*

Part I: Let $f(x)$ *be upward sloping, i.e.,* $f'(x) > 0 \; \forall x \in [0, 1]$*.*

- *(i) There always exists at least one stable equilibrium.*
- *(ii) Sufficient conditions for a unique stable equilibrium are either* $|\Delta V'(x)| \leq 4/\beta \; \forall x \in [0, 1]$ *or* $x^F \notin [0, 1]$*.*
- *(iii) If there is a fixed point* $x^* \in [0, 1]$ *with* $f'(x^*) > 1$*, then* x^* *is not stable, but there exist two stable equilibria, a brown equilibrium with* $x^* = x_b^* < 1/2$ *and green equilibrium with* $x^* = x_g^* > 1/2$*.*

Part II: Let $f(x)$ *be downward sloping, i.e.,* $f'(x) < 0 \; \forall x \in [0, 1]$*.*

- *(iv) There exists a unique equilibrium, though it is not necessarily stable.*
- *(v) If* $f'(x) \geq -1 \; \forall x \in [0, 1]$*, then there exists a unique stable equilibrium.*
- *(vi) If* $f'(x^*) < -1$*, there is no stable equilibrium and* x_t *converges to a period-2 orbit.*[3]

Proof. Follows from the discussion above. Further details are provided in Appendix A. □

The distribution of choices $f(x)$ depends crucially on the intensity of choice β. The lower β, the flatter $f(x)$, with $f(x)$ being a horizontal line in the limit $\beta \to 0$ (Figure 2, panel (a)). Then $x^* = \lim_{\beta \to 0} f(x) = 1/2$ is the unique stable equilibrium. In this case, the two lifestyles have the same probability of being adopted, and the population splits equally in equilibrium. In contrast, the larger the value of β, the more pronounced the S-shape of the distribution $f(x)$, with $f(x)$ becoming a step function in the limit $\beta \to \infty$. Different possible step functions are shown in Figure 2 for $\beta \to \infty$.

Panels (b) (c) and (d) in Figure 2 show an upward sloping function $f(x)$. In panel (b) and (d) the flex point is outside the domain of x. Therefore, we have one stable equilibrium at the boundary of x. In panel (b), $x^F < 0$ and $\Delta V(x) < 0$ for all $x \in [0, 1]$. Hence, all individuals choose the green lifestyle ($x_g^* = x^* = 1$). In panel (d), $x^F > 0$ and $\Delta V(x) > 0$ for all $x \in [0, 1]$. Hence, all individuals choose the brown lifestyle ($x_b^* = x^* = 0$). In panel (c) the flex point is inside the domain of x, i.e., $x^F \in [0, 1]$. Hence, there are two extreme stable equilibria, $x_b^* = x^* = 0$ and $x_g^* = x^* = 1$, and one unstable equilibrium in between.

---

[3] Notice that if $f(x)$ is downward sloping, the sufficient condition for a stable equilibrium also includes $f'(x) = -1$. Instead, if $f(x)$ is upward sloping, a fixed point with $f'(x^*) = 1$ is unstable if it is a tangent bifurcation point, while it is stable if it is a pitchfork bifurcation point, as this will be explained in more detail in the following sections. The condition $f'(x) < 1 \; \forall x \in [0, 1]$ does not allow for bifurcation values.

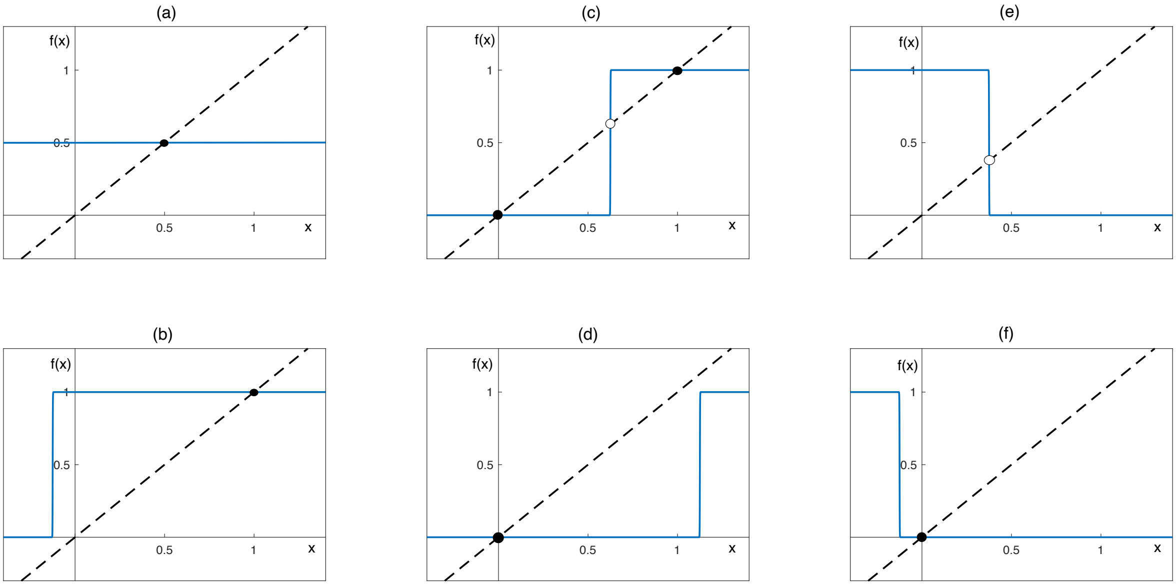

Figure 2: Distribution of $f(x)$ for extreme values of β. In case (a), $\beta = 0$. In all other cases, $\beta \to \infty$. For cases (c) and (e), $x^F \in [0, 1]$, with $f(x)$ upward (case (c)) and downward (case (e)) sloping, respectively. In the bottom panels, $x^F \notin [0, 1]$, and $f(x)$ is upward sloping in cases (b) and (d), while it is downward sloping in case (f).

Panels (e) and (f) in Figure 2 show a downward sloping function $f(x)$. For $\beta \to \infty$, we have only an unstable equilibrium in panel (e) because the flex point is inside the domain of x, i.e., $x^F \in [0, 1]$. However, if the flex point is outside this domain of x, $x^F \notin [0, 1]$, like in panel (f), then one of the two extreme equilibria emerges. In panel (f), the case is illustrated where $x_b^* = x^* = 0$.

The following proposition summarises the equilibrium scenarios for extreme values of the intensity of choice β.

Proposition 2.3. *For extreme values of the intensity of choice β the following holds:*

- $\beta \to 0$*: there is one unique stable equilibrium at* $x^* = \frac{1}{2}$.
- $\beta \to \infty$*: (i) if* $x^F \notin [0, 1]$*, then there is a unique equilibrium which is either a brown equilibrium* $x_b^* = x^* = 0$ *or a green equilibrium* $x_g^* = x^* = 1$*; (ii) if* $x^F \in [0, 1]$*, there is no stable equilibrium if* $f(x)$ *is downward sloping and there are two stable equilibria* $x_b^* = x^* = 0$ *and* $x_g^* = x^* = 1$ *if* $f(x)$ *is upward sloping.*

Proof. If $\beta \to 0$, irrespective of ΔV, we have $e^{\beta \Delta V(x)} \to 1$ and $f(x) = \frac{1}{1+e^{\beta \Delta V(x)}} \to 1/2$. The value of $\lim_{\beta \to \infty} e^{\beta \Delta V(x)}$ depends on the sign of ΔV. If $\Delta V > 0$, we have $\lim_{\beta \to \infty} e^{\beta \Delta V(x)} = \infty$ and $f(x') = \frac{1}{1+e^{\beta \Delta V(x)}} \to 0$. If $\Delta V < 0$, we have $\lim_{\beta \to \infty} e^{\beta \Delta V(x)} = 0$, such that $f(x) = \frac{1}{1+e^{\beta \Delta V(x)}} \to 1$. If $\Delta V(x) = 0$, then $f(x) = 1/2$ irrespective of β. □

Note that $\beta \to \infty$ allows us to relate our discrete choice model with random payoff to standard economic models with deterministic payoff. If individuals are perfectly rational or perfectly homogeneous, their decisions only depend on the deterministic part of the payoff. Hence, whenever $\Delta V(x) > 0$, all individuals would choose a brown lifestyle and whenever $\Delta V(x) < 0$, the entire population would choose a green lifestyle. If we have a positive feedback, and the flex point is in $[0, 1]$, like in panel (c) of Figure 2, the brown equilibrium would emerge if the starting value x_0 at time $t = 0$ is to the left of the unstable equilibrium, and the green equilibrium would emerge if the starting value is to the right of the unstable equilibrium. That is, the unstable equilibrium separates the basins of attraction of the two equilibria. If we have a negative feedback and the flex point is in $[0, 1]$, the unstable equilibrium implies an oscillation between two extreme points. Such phenomenon will be discussed in Section 4 in more detail.

We conclude this section with an introduction to social tipping points. A social tipping point is an equilibrium value of $x^* = x^T$, associated with critical parameter values, such that a marginal change of some parameter values causes this equilibrium to disappear. If this equilibrium was populated, the new equilibrium will be very different.

Theorem 1. *Consider the model in equation (7) with the revision protocol* $\hat{x} = x_{t-1}$*, let* $f(x)$ *be upward sloping and* $\mathbf{k}$ *be a vector of parameters values. For a stable equilibrium of the model* x^**, a set of parameters values* $\mathbf{k} = \mathbf{k}_b$ *can always be found such that*

$$\lim_{\mathbf{k}\to\mathbf{k}_b} \frac{\partial x^*}{\partial \mathbf{k}}(x; \mathbf{k}) = \pm\infty. \tag{11}$$

Proof. The difference in payoffs $\Delta V(x, \mathbf{k})$ is linear and the map f(x) in (7) is S-shaped. Consequently, it is always possible to find a parameter configuration where a saddle-node bifurcation at $f'(x^*) = 1$ occurs. This entails a transition from two stable equilibria to one stable equilibrium such that $\left|\lim_{\mathbf{k}\to\mathbf{k}_b} \frac{\partial f}{\partial \mathbf{k}}(x^*; \mathbf{k})\right| = \infty$. □

Theorem 1 asserts that with a positive feedback there always exist parameter values implying a social tipping point. Figure 3 illustrates this result for the cases of tipping points generated by the tangent bifurcation of an upward sloping map $f(x)$, based on the basic model discussed in section 3.[4]

The left panel assumes that we start from the blue function $f(x)$, with a brown ($x_b^* < 1/2$) and green ($x_g^* > 1/2$) stable equilibrium. Then the function shifts to left due to a change of some parameter. Once the red function $f(x)$ is reached, the brown equilibrium becomes unstable and for any further shift to the left of $f(x)$, the brown equilibrium entirely disappears. The bifurcation value in terms of x where this happens is denoted by x_1^T. By the same token, a shift from the blue function $f(x)$ to the right implies that once the purple function $f(x)$ is reached, with the bifurcation value x_2^T, the

---

[4] Tangent bifurcations belong to the one-dimensional version of the more general concept of saddle-node bifurcations (Kuznetsov, 1998).

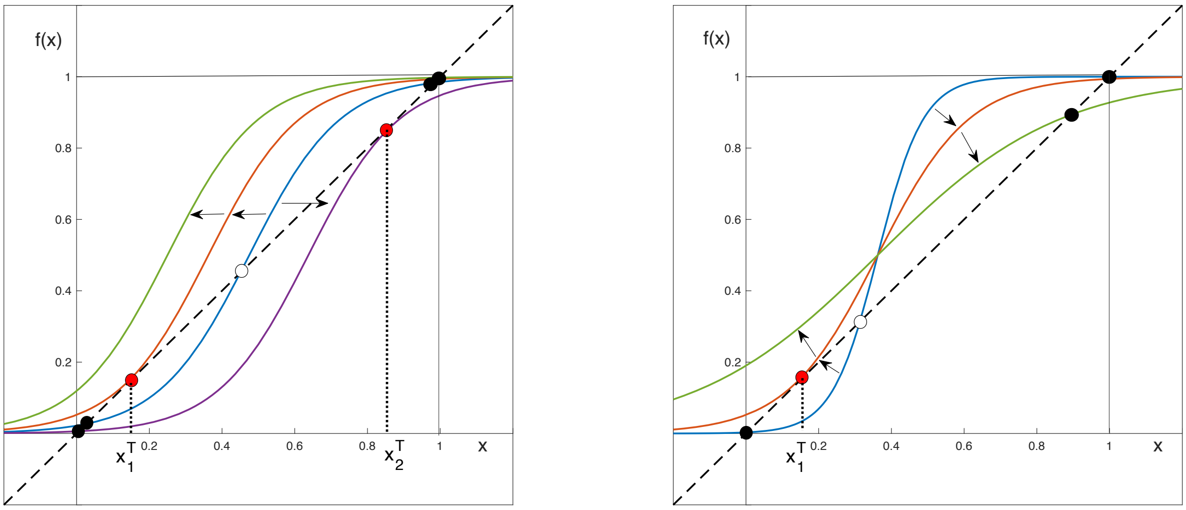

Figure 3: Illustration of tipping points for two examples of tangent bifurcations. The example in the left panel illustrates tipping points (x_1^T and x_2^T) from a shift of the map f due to change of parameter λ, which captures a portion of the relative payoff difference between browns and greens in the basic model. (For further details, see section 3.) The example in the right panel illustrates a tipping point (x_1^T) from a change in the parameter β. The examples are generated with a basic model in section 3 with a logistic function $f(x) = \frac{1}{1+e^{\beta[\lambda+\rho(1-2x)]}}$, where $\rho = 2$, and in the left panel, we have $\beta = 2$ with $\lambda = -0.1$ (blue), $\lambda = -0.55$ (red), $\lambda = -1$ (green), $\lambda = 0.55$ (purple), while in the right panel we have $\lambda = -0.55$ and $\beta = 4$ (blue), $\beta = 2$ (red) and $\beta = 1$ (green).

green equilibrium becomes unstable and any further shift to the right makes the green equilibrium to disappear.

The right panel of Figure 3 illustrates the tipping point induced by a change of parameter β. By lowering β, the blue function $f(x)$ moves gradually to the red and finally to the green function $f(x)$. This implies that the brown equilibrium becomes unstable for a value of β which induces the equilibrium $x_b^* = x_1^T$. Lowering β further will make the brown equilibrium to disappear.

More generally, both panels illustrate that a marginal change of a parameter may imply the disappearance of an equilibrium and a large shift of equilibrium behaviour at a tipping point. Important for our definition is that the new equilibrium already existed, but was not populated. Technically, at the tipping point, the marginal effect of a parameter on the equilibrium value x^* becomes infinite. Note that in our model x^* is a function of parameters. Hence, tipping points x^T are the result of parameter changes. (That is, it would be wrong to claim that at a tipping point a critical fraction of the population just requires some few more supporters in order for the equilibrium to tip.)

In the remainder of the paper, we consider different versions of this general model. We start with a basic model in Section 3, which studies the effect of personal and social norms on equilibrium behaviour under standard assumptions. In Section 4 and 5, we analyse some interesting alternative assumptions about social norms and psychological well-being.

3 The Basic Model

3.1 Ingredients

In order to derive further interesting results, we need to be slightly more specific about the different components in the payoff function of individuals in eq. (1). For simplicity, we assume a quadratic utility function from individual consumption and a linear environmental damage function from total consumption (with $v > 0$ and $\eta > 0$):

$$U_i(q_i) = vq_i - \frac{\eta q_i^2}{2}, \tag{12}$$

$$D(Q) = \delta \sum_{i=1}^{n} q_i = \delta Q. \tag{13}$$

That is, marginal utility is decreasing at the rate η. It will become apparent below that in equilibrium marginal utility will always be positive, i.e., $U'(q_i^*) > 0$ as $q_i^* < \frac{v}{\eta}$. Damages increase at the constant rate $\delta > 0$. Moreover, we consider well-being derived from choosing a green lifestyle to be given by a constant 'warm-glow' parameter $\gamma > 0$. Moreover, social utility, which an individual derives from belonging to a particular group, increases linearly in group size, with parameter $\rho > 0$ capturing the intensity of social interaction.

$$G_i = \begin{cases} \gamma & \text{if green} \\ 0 & \text{if brown} \end{cases} \tag{14}$$

$$I_i = \begin{cases} \rho x & \text{if green} \\ \rho(1-x) & \text{if brown.} \end{cases} \tag{15}$$

Thus, one channel of spillovers in individuals' payoffs is due to the social utility term where proportional social spillovers is a certainly a simplification but customary in the literature on (Brock and Durlauf, 2001). A further channel of spillovers works through utility of material consumption, as also prices are function of x. This is evident below.

3.2 The Constituent Game at Time t

At a particular time t, consider the constituent game with two stages. Given the choice of lifestyle in stage 1 (i.e., the fraction of x is given), individuals maximise their payoff in stage 2 by choosing their consumption levels. In order to motivate that greens and browns choose different consumption levels, we assume that browns choose their consumption level by evaluating marginal damages with parameter δ_b and greens by parameter δ_g, with $0 \leq \delta_b < \delta_g$. In this simple specification, neither the term G_i nor the term I_i in the payoff function are a function of quantities and the maximisation of payoff leads to the first order condition in an interior equilibrium with marginal utility minus price being set equal to marginal damages, $v - \eta q_i - p(x) = \delta_i$. This leads to the following equilibrium quantities, denoted by q_b and q_g, respectively:

$$q_b^*(p(x)) = \frac{v - \delta_b - p(x)}{\eta} \quad \text{and} \quad q_g^*(p(x)) = \frac{v - \delta_g - p(x)}{\eta}, \tag{16}$$

with $q_b^* > q_g^*$ due to our assumption $\delta_b < \delta_g$. Total demand is given by

$$\begin{aligned} Q^*(p(x)) &= n\Big[x q_g^*(p(x)) + (1-x) q_b^*(p(x))\Big] \\ &= \frac{n}{\eta}\left[v - p(x) - \delta_b - (\delta_g - \delta_b)x\right]. \end{aligned} \tag{17}$$

We assume a linear supply function:

$$S(p) = sp \tag{18}$$

where s is a parameter. In the market equilibrium, we have $Q^*(p^*(x)) = sp^*(x)$. Solving for the equilibrium price gives:

$$p^*(x) = \frac{v - \delta_b - (\delta_g - \delta_b)x}{1 + \eta\frac{s}{n}}. \tag{19}$$

By substituting (19) into (17) and (16), we obtain total and individual equilibrium consumption $Q^*(x)$ and $q_i^*(x)$. As $\delta_g > \delta_b$, equilibrium price $p^*(x)$ and total equilibrium consumption $Q^*(x)$ decrease in the fraction of individuals x adopting a green lifestyle in line with Proposition 2.1.[5]

In the first stage, given the consumption level of the second stage, the choice of lifestyle depends on the difference between the brown and green payoff, $\Delta V(x) = V_b(x) - V_g(x)$.

$$V_i = \begin{cases} V_g(x) = U(q_g^*(x)) - p^*(x)q_g^*(x) - \delta Q^*(x) + \gamma + \rho x & \text{if green} \\ V_b(x) = U(q_b^*(x)) - p^*(x)q_b^*(x) - \delta Q^*(x) + \rho(1-x) & \text{if brown.} \end{cases} \tag{20}$$

Hence, after substituting the equilibrium price $p^*(x)$ from (19),

$$\Delta V(x) = \zeta - \gamma + \rho(1 - 2x), \tag{21}$$

with $\zeta = \frac{\delta_g^2 - \delta_b^2}{2\eta}$. The first term of $\Delta V(x)$ in (21) is the difference between the brown and green lifestyle in terms of utility derived from different consumption levels. This (positive) term increases in the difference between the degree of internalisation of the consumption externality of greens and browns captured by the parameters δ_g and δ_b. This term decreases in the slope of the marginal utility function, captured by the parameter η. The second term γ captures the disadvantage of browns compared to greens, as they do not

---

[5] We assume the conditions for a strictly positive price and quantities to hold.

derive a warm glow from a green lifestyle. The third term captures the difference in social utility derived by browns and greens.[6] We notice that $\Delta V(x)$ is linear in the fraction x of individuals adopting a green lifestyle in line with the assumption of Proposition 2.2. In order to write $\Delta V(x)$ in a more compact form, we define $\lambda = \zeta - \gamma$. Hence, $\Delta V(x) = \lambda + \rho(1-2x)$. The parameter λ measures the relative attractiveness of the two lifestyles, aggregating utility derived from consumption and warm glow. The value of x that makes agents indifferent between the two lifestyles ($\Delta V(x^F) = 0$) is given by

$$x^F = \frac{1}{2}\left(1 + \frac{\lambda}{\rho}\right), \tag{22}$$

where we recall from Section 2 that x^F is the flex point of $f(x)$ where $f''(x^F) = 0$.

Proposition 3.1. *In the basic model, the indifference point x^F is characterised as follows.*

(i) x^F increases in λ; x^F decreases in ρ if $\lambda > 0$, increases in ρ if $\lambda < 0$ and remains constant and equal to $\frac{1}{2}$ if $\lambda = 0$.

(ii) Let $\lambda < -\rho \leq 0$, then $x^F \notin [0,1]$, with $x^F < 0$.

(iii) Let $-\rho \leq \lambda \leq 0$, then $0 \leq x^F \leq \frac{1}{2}$.

(iv) Let $0 < \lambda \leq \rho$, then $\frac{1}{2} < x^F \leq 1$.

(v) Let $0 \leq \rho < \lambda$, then $x^F \notin [0,1]$, with $x^F > 1$.

Proof. Follows from basic calculations using (22). □

Notice that λ can take on negative and positive values, whereas ρ is always strictly positive. We observe that the larger λ, the more attractive is the brown compared to the green lifestyle. Hence, the flex point increases, i.e., a larger x is needed such that green consumers receive sufficient social utility from choosing a green lifestyle in order to make both lifestyles equally attractive. Moreover, the impact of the social interaction parameter ρ on the indifference point x^F depends on the sign of λ. If the parameter λ takes on values above or below a threshold defined with respect to the social interaction parameter ρ, the indifference point is outside the domain $[0,1]$, otherwise it is inside this domain. If it is outside, either $\Delta V(x) < 0$ or $\Delta V(x) > 0$ for all possible values of x.

---

[6] We assume on purpose that the payoff functions of all individuals are identical such that different payoffs are just due to different lifestyles associated with different behaviours. If we assumed the evaluation of environmental damages to be different between browns and greens, all qualitative results would continue to hold. That is, $\Delta V(x)$ is linearly decreasing in x and therefore the function $f(x)$ is upward sloping over the entire domain of x.

3.3 Equilibrium Analysis

The equilibrium value of the fraction of green choices is a fixed point of the logistic distribution $f(x)$ of eq. (7) which depends on the discrete choice utility difference $\Delta V(x) = \lambda + \rho(1-2x)$. The slope of this logistic function, $f'(x)$, follows from eq. (8). Noticing that $\Delta V'(x) = -2\rho$, this gives $f'(x) > 0$ for all values of x. Consequently, there is a positive feedback. Therefore, Part I in Proposition 2.2 applies: there is either one equilibrium, which is also stable, or three equilibria of which two are stable. Moreover, we know that a sufficient condition for a unique equilibrium is if $f'(x^F) < 1$. Using eq. (10) and the fact that $\Delta V'(x) = -2\rho$, we have $f'(x) < 1 \Rightarrow \beta\rho < 2$. We also know that another sufficient condition for a unique equilibrium is if the indifference point is outside the domain [0,1], which, using the information in Proposition 3.1, gives $\lambda < -\rho < 0$ and $0 < \rho < \lambda$. For convenience, Corollary 3.1 summarizes these results.

Corollary 3.1. *In the basic model, we have:*

(i) a positive feedback for the variable x_t;

(ii) one stable equilibrium or two stable equilibria of x_t;

(iii) sufficient conditions for a unique stable equilibrium are

- *a. $\beta\rho \leq 2$,*
- *b. $\lambda < -\rho \leq 0$,*
- *c. $0 \leq \rho < \lambda$.*

The results of Corollary 3.1, part (iii), are illustrated in Figure 4 where we also indicate the complementary necessary conditions for two stable equilibria.

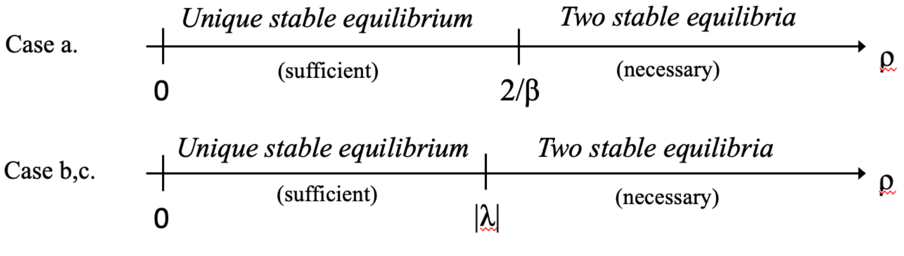

Figure 4: Different equilibrium scenarios in the social interaction intensity space ρ according to Corollary 3.1.

The sufficient conditions for a unique stable equilibrium, expressed in terms of the social interaction parameter ρ, imply that only if the social interaction is not too strong, can one of the three sufficient conditions for a unique equilibrium be satisfied. Examples of a unique stable equilibrium are given in Figure 5, in the left and centre panel. A similar conclusion emerges in terms of the parameter β, the intensity of choices if the

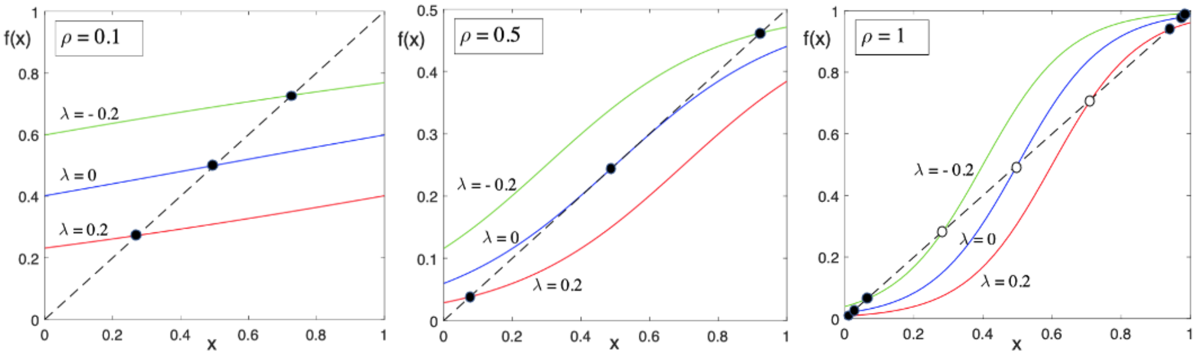

Figure 5: Map $f(x)$ for different values of parameter λ and ρ in the basic model. Left: $\rho = 0.1$. Centre: $\rho = 0.5$. Right: $\rho = 1$. $\beta = 4$ is assumed throughout. Circled dots indicate unstable, filled dots stable equilibria.

flex point is inside the domain of x. Only if β is sufficiently small will there be a unique equilibrium. Conversely, if the degree of social interaction ρ is strong (or the intensity of choices β is high), then there will be two stable equilibria, a brown with $x_b^* < 1/2$ and a green equilibrium with $x_g^* > 1/2$. An example is given in the right panel in Figure 5. In fact, as we will analyse in more detail in section 3.4, the larger ρ and β, the smaller will be the equilibrium x_b^* and the larger will be the equilibrium x_g^*. In other words, x_b^* moves towards zero and x_g^* moves towards one, if either ρ or β increases. That is, the two equilibria move to the boundaries of the domain x if either ρ or β increase and if both values are large. This is evident from Figure 5 by comparing $f(x)$ for a given value of λ across the three panels. This conclusion is reconfirmed if we consider the limit when β or ρ become very large.

Corollary 3.2. *In the basic model, the following holds:*
Part I: if $\beta \to \infty$, then

(i) $x_b^* = 0$ is the unique equilibrium if $\rho < \lambda$ and $\lambda > 0$,

(ii) $x_g^* = 1$ is the unique equilibrium if $\rho < -\lambda$ and $\lambda < 0$,

(iii) $x_b^* = 0$ and $x_g^* = 1$ are the two stable equilibria if $-\rho \leq \lambda \leq \rho$;

Part II: if $\rho \to \infty$, then $x_b^* = 0$ and $x_g^* = 1$ are the two stable equilibria and $x^F = 1/2$.

Proof. Part I follows directly from the general insights of Proposition 2.3 in Section 2. Part II follows, as in the limit for $\rho \to \infty$, $f(x)$ is a step function with a discontinuity exactly at $x^F = \frac{1}{2}$, as $\lim_{\rho\to\infty} x^F = \lim_{\rho\to\infty} \frac{\lambda+\rho}{2\rho} = \frac{1}{2}$. This implies that $f(x^*) = 0$ for $x < \frac{1}{2}$, and $f(x^*) = 1$ for $x > \frac{1}{2}$. Therefore, there are two stable equilibria $x_b^* = 0$ and $x_g^* = 1$. $\square$

That is, strong social interaction and a high intensity of choice favour extreme equilibria at the boundary of the domain of x.

3.4 Parameter Changes and Social Tipping Points

In this subsection, we analyse how the most important parameters of our basic model affect the equilibrium value of the state variable x_t. In particular, we devote attention to parameter bifurcations, which, as anticipated in Theorem 1 in Section 2, are critical values at which an equilibrium disappears, i.e., social tipping points. We recall that β measures the intensity of lifestyle choices, ρ measures the intensity of social interaction and λ is a 'summary parameter', $\lambda = \frac{\delta_g^2 - \delta_b^2}{2\eta} - \gamma$, which measures the relative attractiveness of the brown over the green lifestyle from material consumption and warm-glow.

Lemma 3.1. *The effect of λ on the stable equilibrium value $x^* = f(x^*)$ is given by*

$$\frac{\partial x^*}{\partial \lambda} = \frac{\beta x^*(1 - x^*)}{2\rho\beta x^*(1 - x^*) - 1}. \tag{23}$$

The following cases in terms of the sign of (23) can be distinguished, considering an interior equilibrium $0 < x^* < 1$.

Part I: If $\rho\beta \leq 2$, then $\frac{\partial x^*}{\partial \lambda} \leq 0$ in the unique stable equilibrium x^*.

Part II If $\rho\beta > 2$, and if two stable equilibria exist, two characteristic values, λ_1^T and λ_2^T, associate with the social tipping points x_1^T and x_2^T exist, such that

- *(i)* $\frac{\partial x^*}{\partial \lambda} < 0$ *in a stable brown equilibrium* $0 < x^* = x_b^* < x_1^T < \frac{1}{2}$ *and in a stable green equilibrium* $1 > x^* = x_g^* > x_2^T > \frac{1}{2}$;
- *(ii)* $\lim_{\lambda \to \lambda_2^T} \frac{\partial x^*}{\partial \lambda} = -\infty$ *with* $\lim_{\lambda \to \lambda_2^T} x^* = x_2^T > \frac{1}{2}$;
- *(iii)* $\lim_{\kappa \to \kappa_1^T} \frac{\partial x^*}{\partial \kappa} = +\infty$ *with* $\kappa = -\lambda$, $\kappa_1^T = -\lambda_1^T$ *and* $\lim_{\kappa \to \kappa_1^T} x^* = x_1^T < \frac{1}{2}$,

where

$$x_1^T = \frac{1 - \sqrt{1 - \frac{2}{\rho\beta}}}{2}, \quad x_2^T = \frac{1 + \sqrt{1 - \frac{2}{\rho\beta}}}{2} = 1 - x_1^T, \tag{24}$$

and

$$\lambda_1^T = \frac{1}{\beta} \ln\left(\frac{1 + \sqrt{1 - \frac{2}{\rho\beta}}}{1 - \sqrt{1 - \frac{2}{\rho\beta}}}\right) - \rho\sqrt{1 - \frac{2}{\rho\beta}} \tag{25}$$

$$\lambda_2^T = \frac{1}{\beta} \ln\left(\frac{1 - \sqrt{1 - \frac{2}{\rho\beta}}}{1 + \sqrt{1 - \frac{2}{\rho\beta}}}\right) + \rho\sqrt{1 - \frac{2}{\rho\beta}} = -\lambda_1^T. \tag{26}$$

Proof. See Appendix B, assuming ρ and β to be finite. □

Given the parameters contained in λ, their effects on x^* follow immediately, which is summarised in the following proposition.

Proposition 3.2. *The effect of parameters δ_g, δ_b and γ on an interior stable equilibrium $x^* = f(x^*)$ is given by*

- $\frac{\partial x^*}{\partial \delta_g} = \frac{\partial x^*}{\partial \lambda}\frac{\partial \lambda}{\partial \delta_g} = \frac{\delta_g}{\eta}\frac{\partial x^*}{\partial \lambda}$;
- $\frac{\partial x^*}{\partial \delta_b} = \frac{\partial x^*}{\partial \lambda}\frac{\partial \lambda}{\partial \delta_b} = -\frac{\delta_b}{\eta}\frac{\partial x^*}{\partial \lambda}$;
- $\frac{\partial x^*}{\partial \gamma} = \frac{\partial x^*}{\partial \lambda}\frac{\partial \lambda}{\partial \gamma} = -\frac{\partial x^*}{\partial \lambda}$.

Both in the unique equilibrium regime ($\rho\beta \leq 2$) and in the two-equilibria regime ($\rho\beta > 2$), we have $\frac{\partial x^*}{\partial \delta_b} > 0$, $\frac{\partial x^*}{\partial \gamma} > 0$ and $\frac{\partial x^*}{\partial \delta_g} < 0$. In the two-equilibria regime, at the tipping point x_1^T, a marginal increase of δ_b or γ causes the brown equilibrium to disappear, and at the tipping point x_2^T, a marginal increase of δ_g causes the green equilibrium to disappear.

Part I of Lemma 3.1 refers to the case where $f(x)$ is relatively flat. Therefore, we have a unique stable equilibrium. Regardless of where the equilibrium is located, if λ increases (decreases), the brown (green) lifestyle becomes relatively more attractive and the fraction of individuals adopting the green lifestyle decreases (increases). This result is illustrated in the left and central panel in Figure 5.

The same direction of change is true for Part II of Lemma 3.1, result (i), though now the function $f(x)$ is sufficiently steep such that we can have two stable equilibria, a brown and a green equilibrium. Result (i) is illustrated in Figure 5 in the right panel. Increasing λ moves $f(x)$ to the right and lowers x^* and decreasing λ moves $f(x)$ to the left and increases x^*.

Part II, result (ii), in Lemma 3.1 captures the idea that λ is at the bifurcation level λ_2^T with fraction x_2^T such that any marginal increase of λ, say through an increase of δ_g, renders the green equilibrium to disappear. This is illustrated in Figure 3 in the left panel. The purple function $f(x)$ is tangent to the 45^o-line. Any further increase of λ would move $f(x)$ to the right and the green equilibrium would disappear. By the same token, Part II, result (iii), in Lemma 3.1 captures the mirror image: λ is small and at the bifurcation level λ_1^T with fraction x_1^T such that any marginal decrease of λ, say through an increase of δ_b or γ, renders the brown equilibrium to disappear. This is shown in Figure 3 in the left panel: the red function $f(x)$ is tangent to the 45^o-line. Lowering marginally λ makes the brown equilibrium to disappear.

Proposition 3.2 is a direct application of Lemma 3.1 for some important parameters of the basic model. Increasing the awareness of brown consumers about the external costs of their consumption (i.e., an increase of δ_b) and/or an increase of the warm glow which green consumers derive (i.e., an increase of γ) increases the equilibrium fraction of greens in the population and may transform a society abruptly by jumping from a brown to a green equilibrium at a social tipping point. The reverse is true if the awareness of an already green population is increased. An already "green society" may jump to a brown equilibrium at a social tipping point. Thus, raising the environmental awareness of the population about the negative effects of consumption should focus on brown and not on green consumers. Turning brown into green consumers increases x, turning greens into "supergreens" does the opposite. This may quite relevant as very often those with higher environmental awareness are those being more responsive to environmental campaigns.

Whereas at a social tipping, marginal changes of parameters imply abrupt and non-continuous changes of an equilibrium, once a threshold has been passed, further, even non-marginal changes of parameters, have only small equilibrium effects. That is, once an equilibrium has tipped, this change is very robust to pertubations and cannot easily be reversed. This is explained with the help of Figure 6.

The two continuous curves represent the loci of stable equilibria for different values of λ. For the upper curve, these are green equilibria, for the lower curve, they are brown equilibria. The middle dashed leg of the curve are unstable equilibria, with the small arrows indicating the basins of attraction of the stable equilibria. The long dashed arrows indicate the jump occurring at a social tipping point, x_1^T and x_2^T, occurring at the bifurcation values λ_1^T and λ_2^T, respectively.

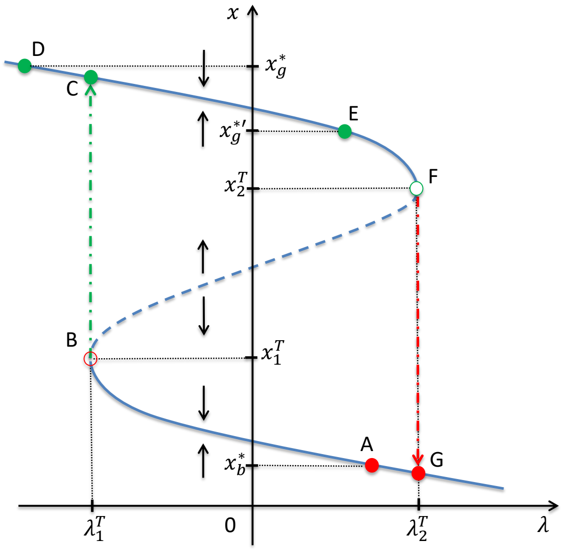

Figure 6: Representation of long-run equilibria of x (vertical axis) for different values of the parameter λ (horizontal axis) in a scenario with multiple equilibria. The solid lower and upper curves are loci of stable equilibria, while the dashed curve is the loci of unstable equilibria. Black arrows represent field lines of the basins of attraction for x. Bifurcation values λ_1^T and λ_2^T are indicated, as well as the corresponding tipping points x_1^T and x_2^T. The points from A to G exemplify a 'hysteresis' cycle from a reduction and then an increase of λ. Coloured dashed arrows indicate tipping events: green arrow with a tipping from a brown to a green equilibrium, and red with a tipping from a green to a brwon equilibrium.

Suppose β and ρ are sufficiently large and that a green and a brown equilibrium exist. Assume initially a positive and large value of λ and that the population is in brown equilibrium x_b^* (point A). Consider now that λ gradually decreases (e.g., because the warm glow which greens derive increases over time and/or the environmental awareness of brown consumers increases), then the fraction of individuals adopting a green lifestyle will increase. This continues until λ reaches the value λ_1^T where any further marginal decrease of λ will tip the equilibrium from a brown ($x_b^* = x_1^T$) to a green equilibrium ($x_g^* > x_2^T$), i.e., there is a jump from point B to point C. Decreasing λ further makes x_g^* larger, but x_g^* changes only marginally, such that for instance point D is reached.

Starting from point D, consider now a change of λ in the opposite direction, say because the awareness of green consumers increases. Continuously increasing λ will only lead to a marginal decrease of fraction x where the population is in a green equilibrium. Even if λ passes the bifurcation value λ_1^T, the change in x_g^* continues to be marginal, as λ_1^T is the tipping point of a brown equilibrium, but not of a green equilibrium. Only if λ increases until it reaches the bifurcation value λ_2^T will the green equilibrium disappear, i.e., $x_g^* = x_2^T$ disappears and the population tips to a brown equilibrium ($x_b^* < x_1^T$), i.e., there is a jump from point F to point G. Thus, once an equilibrium tips, it will not immediately tip back just because of small parameter changes. Thus, sudden social transformation are robust.

Note that gradual changes but also jumps at social tipping points can also be induced by parameters β and ρ.

Proposition 3.3. *The effect of β and ρ on a stable equilibrium x^* is given by*

$$\frac{\partial x^*}{\partial \beta} = \frac{x^*(1-x^*)\Delta V}{2\rho\beta x^*(1-x^*)-1} = \frac{\Delta V}{\beta}\frac{\partial x^*}{\partial \lambda}; \tag{27}$$

$$\frac{\partial x^*}{\partial \rho} = \frac{\beta x^*(1-x^*)(1-2x^*)}{2\rho\beta x^*(1-x^*)-1} = (1-2x^*)\frac{\partial x^*}{\partial \lambda}. \tag{28}$$

Let $\phi = \beta, \rho$. The following cases in terms of the sign of (27) and (28) can arise, considering an interior equilibrium $0 < x^* < 1$.
In the unique and two-equilibria regime, we have:

- $\frac{\partial x^*}{\partial \phi} < 0$ *if* $x^* < \frac{1}{2}$ *(i.e.,* $\Delta V(x^*) > 0$*);*
- $\frac{\partial x^*}{\partial \phi} \geq 0$ *if* $x^* \geq \frac{1}{2}$ *(i.e.,* $\Delta V(x^*) \leq 0$*).*

In the two-equilibria regime, a critical value ϕ^T exists such that, for $\psi = -\phi$ and $\psi^T = -\phi^T$, we have:

$\lim_{\psi\to\psi^T} \frac{\partial x^*}{\partial \psi} = -\infty$ *with* $\lim_{\psi\to\psi^T} x^* = x_2^T > \frac{1}{2}$*;*

$\lim_{\psi\to\psi^T} \frac{\partial x^*}{\partial \psi} = +\infty$ *with* $\lim_{\psi\to\psi^T} x^* = x_1^T < \frac{1}{2}$*.*

Proof. See Appendix B. □

Proposition 3.3 is illustrated in the right panel of Figure 3 for the parameter β. Increasing β makes the function $f(x)$ steeper around the flex point and flatter at the edges. This "reinforces" equilibrium values. x^* decreases if it is a brown equilibrium ($x_b^* < \frac{1}{2}$) and x^* increases if it is a green equilibrium ($x_g^* > \frac{1}{2}$), irrespective whether there is a unique stable equilibrium or two stable equilibria. However, if there are two stable equilibria, at the tipping points x_1^T and x_2^T, an equilibrium suddenly disappears, depending on the direction of change of this parameter. In Figure 3, the case is illustrated where at the tipping point, the red curve for $f(x)$ is tangent to the 45^o-line and where any marginal decrease of β makes the brown equilibrium to disappear. Similarly, though not illustrated in Figure 3, if we were in a green equilibrium, any marginal reduction of β would make the green equilibrium to disappear. For the parameter ρ, almost same holds, which is illustrated in Figure 5. Moving from right to left panel, the green equilibrium disappears for the red curve $f(x)$ and the brown equilibrium for the green curve $f(x)$. The logic is that large values of β and ρ reinforces equilibria as pointed above. Therefore, the opposite is needed to tip an equilibrium. Hence, a society may be trapped in a brown equilibrium if social interaction is strong and if the homogeneity among consumers'preferences is high. In this case, the transformation of society requires a substantial increase of the awareness of brown consumers about the externality they cause through consumption and/or a very large increase of the well-being derived from green lifestyle.

4 Extension I: Non-linear Social Norm

4.1 Non-linear Social Utility

In the basic model, we assume that individuals derive social utility from belonging to a group who have adopted the same lifestyle. The larger the group to which an individual belongs, the larger is social utility an individual derives. This could be the case for instance because individuals perceive that their behaviour is in line with the majority in their society, which they may view as a social norm and which confirms them of doing the "right thing". This incentivises individuals to follow the crowd. However, even though individuals may want to be confirmed in their behaviour, and not being the only one adopting a particular lifestyle, they may also not like being perceived as following the mainstream. Thus, social utility may have some features of crowding-out; it captures the role of status in decision-making, or what is called the "minority influence effect" in social psychology.

In order to capture this phenomenon, we assume that social utility of greens only increases (at a decreasing rate) if the fraction of individuals adopting a green lifestyle x

is below a threshold, but decreases (at an increasing rate) beyond this threshold. Thus, there is some crowding out after a threshold of x has been reached. By the same token, we assume the same for browns for whom social utility increases in the fraction $1-x$. That is, a lifestyle becomes increasingly less attractive if it is too widespread in the population. For concreteness, consider eq. (29) instead of eq. (15) for the social utility term in the payoff function of an individual:

$$I_i = \begin{cases} \rho x(a-x) & \text{if green} \\ \rho(1-x)[a-(1-x)] & \text{if brown.} \end{cases} \quad a \in (0,2]. \tag{29}$$

The function $I_g(x) = \rho x(a-x)$ increases in x for $[0, a/2)$, obtains its maximum at $a/2$, and decreases in x for $(a/2, 1]$, whereas $I_b(x) = \rho(1-x)[a-(1-x)]$ increases in x for $[0,(2-a)/2)$, obtains its maximum at $(2-a)/2$ and decreases in x for $((2-a)/2, 1]$. These two functions are symmetric with respect to $x = \frac{1}{2}$, as $I_g(1/2) = I_b(1/2)$. For $a \in [1,2]$, the crowding-out effect occurs for each group if they represent (weakly) more than 50% of the population. For $a \in (0,1)$, the crowding-out occurs below 50%. Figure 7 shows an example for each of these two scenarios.

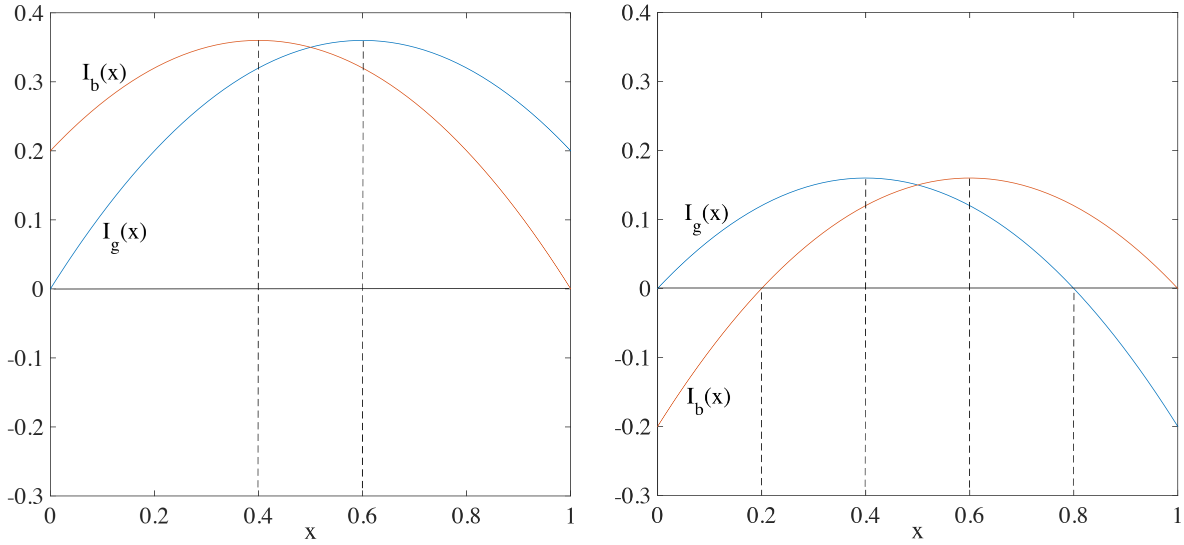

Figure 7: The social utility functions of the green and brown lifestyle, $I_g(x)$ and $I_b(x)$, given in eq. (29). Left: example assumes $a = 1.2$. Right: example assumes $a = 0.8$. In both examples, $\rho = 1$ is assumed.

Note that for $a \in (0,1)$, social utility I_g and I_b can be negative. However, given that not absolute utility, but the difference between utility levels matters for lifestyle decisions, this can be discarded.[7]

Note that equilibrium consumption and price are the same as in the basic model (eq. (16) and eq. (19), respectively).

---

[7] For instance, if we rescale social utility with two constant terms $g, b > 0$, such that $I'_g = I_g + g$ and $I'_b = I_b + b$, and impose that $I_g(x=1) = I_b(x=0) = 0$, we get $b = g$ and these constant terms cancel out in ΔV.

4.2 Stability analysis

Assuming all components of the payoff, except the social utility which is now eq. (29) instead of eq. (15), to be the same as in the basic model, the difference in payoff ΔV between brown and green lifestyles is given by

$$\Delta V(x) = V_b - V_g = \lambda + \rho(a-1)(1-2x). \tag{30}$$

where we recall that $\lambda = \zeta - \gamma = \frac{\delta_g^2 - \delta_b^2}{2\eta} - \gamma$. The revision protocol of this model is again the logistic function $f(x)$ of (7), with $\Delta V(x)$ as given by (30). For $a = 2$, we are back in the basic model. The case $a = 1$ is degenerated, as $I_g = I_b$. Hence, decisions are not influenced by social utility. Henceforth, we do not consider this case.

The value of the state variable x that makes agents indifferent ($\Delta V = 0$) is again the flex point of f, which IS now

$$x^F = \frac{1}{2}\left(1 + \frac{\lambda}{\rho(a-1)}\right). \tag{31}$$

Moreover, $\Delta V'(x) = -2(a-1)\rho$. Thus, the slope of the logistic function $f'(x)$, as given by (8), is positive if $a > 1$ and is negative if $a < 1$.

Corollary 4.1. *For $a > 1$, the revision protocol of the repeated discrete choice for extension I implies a positive feedback and for $a < 1$ a negative feedback of social interaction.*

Since the case of positive feedback leads to the same qualitative conclusions of the basic model discussed in Section 3, we exclusively focus in the remainder of this section on the case of negative feedback. A negative feedback captures the dynamics of a minority game: the more agents adopting a particular lifestyle, the less attractive this lifestyle becomes.

Corollary 4.2. *Let $a < 1$. In the model with non-linear social utility (extension I), a sufficient condition for a unique stable equilibrium is $\rho\beta < \frac{2}{1-a}$.*[8]

Obviously, the opposite of the sufficient condition for a unique stable equilibrium is the necessary condition for an unstable equilibrium. The sufficient and necessary conditions are illustrated in Figure 8 for a negative feedback ($a < 1$). An example of a stable equilibrium is shown in the left panel in Figure 9. In the equilibrium, we have $f'(x^*) > -1$. An example of an unstable equilibrium is shown in the right panel in Figure 9, where $f'(x^*) < -1$ for the fixed point x^*. The system converges to a periodic orbit (Medio and

---

[8] For a negative feedback, a flex point outside $[0,1]$ is not sufficient for a unique stable equilibrium, as this is the case for a positive feedback. For example, assume $\delta_g = 2$, $\delta_b = 0$, $\eta = 1$ (which gives $\zeta = 2$), $\beta = 20$, $\rho = 0.5$, $a = 0.5$, then we have $\rho\beta(1-a) = 5$ and $\rho(1-a) = 0.25$. For $\gamma = 2.26$, we get $\lambda = -0.26$. This implies that the flex point $x^F = 1.02$ is outside the range $[0,1]$. Nevertheless, in this example, the system enters into periodic dynamics.

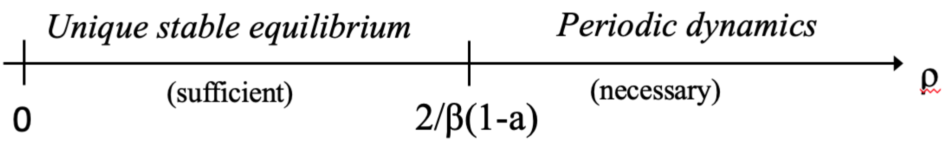

Figure 8: Equilibrium scenarios for the model with non-linear social interaction and negative feedback ($a < 1$).

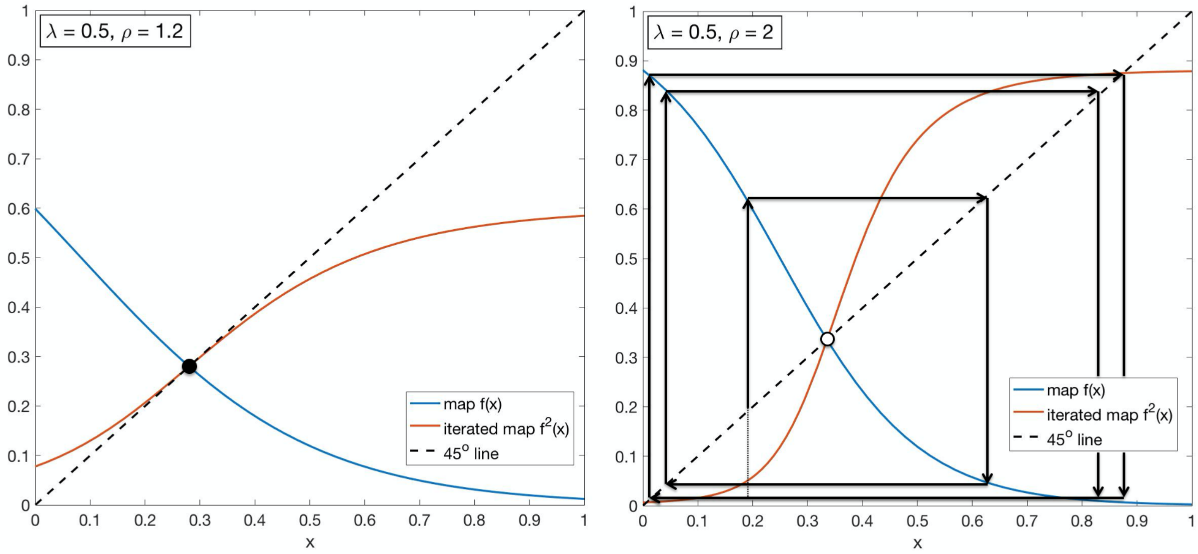

Figure 9: Map f and f^2 of extension I for the negative feedback scenario with $a = 0.5$, $\beta = 4$ and $\lambda = 0.5$ ($\delta_g = 2$, $\delta_b = 0$, $\eta = 1$ and $\gamma = 1.5$). Left: stable equilibrium. Right: periodic orbit.

Gallo, 1995). A small perturbation of the unstable equilibrium x^* from $x^* = 0.35$ to the left, leads for instance to $x = 0.2$. From then onwards, the arrows indicate the revision over time, which lead away from the unstable equilibrium. Essentially, the system enters into periodic dynamics, i.e., an orbit with an oscillation between two values in the long-run, which, in the example, are given by $x_1^{**} \simeq 0.02$ and $x_2^{**} \simeq 0.88$. Figure 10 illustrates these two examples with a time series. The left panel shows a case of convergence to

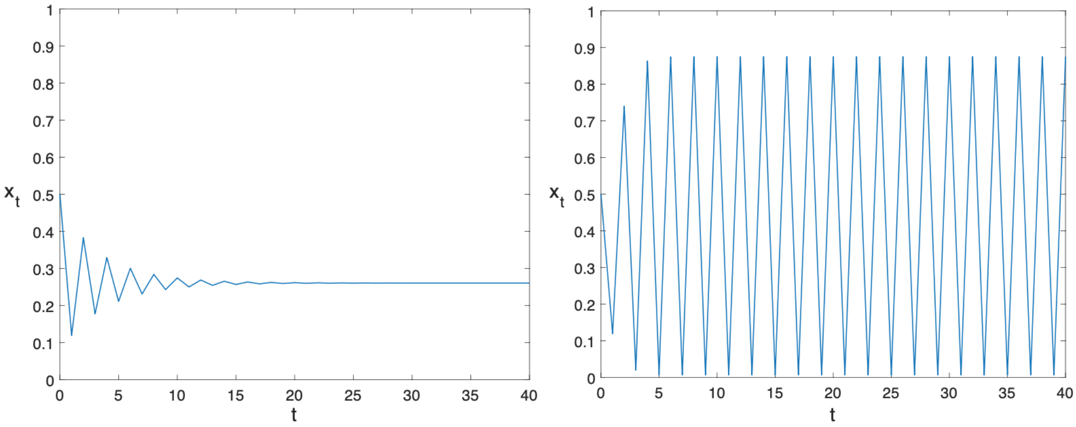

Figure 10: Simulated time series of the fraction of green choices x in the model with non-linear social interaction and negative social feedback, assuming $a = 0.5$, $\beta = 4$ and $\lambda = 0.5$, obtained with $\delta_g = 2$, $\delta_b = 0$, $\eta = 1$ and $\gamma = 1.5$. Left: $\rho = 1$ with convergence to stable equilibrium. Right: $\rho = 2$ with a periodic orbit. In both cases, the initial condition is $x_0 = 0.5$.

the unique stable equilibrium $x^* = 0.25$, the right panel is a case of convergence to a periodic orbit with values $x_1^{**} \simeq 0.02$ and $x_2^{**} \simeq 0.88$. In both cases, the initial value x_0 is irrelevant, as there is a unique attractor, either a stable equilibrium or a periodic orbit.

In order to better understand the difference between stable and unstable equilibria and the amplitude of the oscillation in case of unstable equilibria, we draw on the concept of iterated maps. For the iterated map, we have $f^2(x) = f(f(x)) = f(x) = x$. Thus, the fixed point x^* is also a fixed point of the iterated map and for the two recurring values x_1^{**} and x_2^{**} we have $f^2(x_i^{**}) = f(f(x_i^{**})) = x_i^{**}$ with $i = 1, 2$. Given that $f(x)$ is downward sloping, $f^2(x)$ is upward sloping. We have

$$f^2(x) = \frac{1}{1 + e^{\beta\left[\lambda + \rho(a-1)\left(1 - \frac{2}{1+e^{\beta[\lambda+\rho(a-1)(1-2x)]}}\right)\right]}} \tag{32}$$

with

$$\frac{\partial f^2(x)}{\partial x} = 4\beta^2\rho^2(a-1)^2 \frac{(f^2(x))^2 e^{\beta\left[\lambda + \rho(a-1)\left(1 - \frac{2}{1+e^{\beta[\lambda+\rho(a-1)(1-2x)]}}\right)\right]}}{\left[1 + e^{\beta[\lambda+\rho(a-1)(1-2x)]}\right]^2} > 0. \tag{33}$$

As the map f is S-shaped, the same is true for f^2. If $f(x)$ is downward sloping with a unique stable equilibrium because $f'(x^*) > -1$ (e.g., left panel in Figure 9), the function $f^2(x)$ is upward sloping with unique stable equilibrium exactly at x^*. If $f(x)$ is downward sloping with a unique unstable equilibrium because $f'(x^*) < -1$ (e.g., right panel in Figure 9), the function $f^2(x)$ is upward sloping with an unstable equilibrium exactly at x^* and two stable equilibria x_1^{**} and x_2^{**} which are the long-run equilibria of the original function $f(x)$, which are the two values around which x^* is oscillating.

Proposition 4.1. *The effect of parameters λ, ρ and β on the fixed point $x^* = f(x^*)$ is given by*

$$\frac{\partial x^*}{\partial \lambda} = \frac{\beta x^*(1-x^*)}{-2\rho\beta(1-a)x^*(1-x^*) - 1}, \tag{34}$$

$$\frac{\partial x^*}{\partial \rho} = -(1-a)(1-2x^*)\frac{\partial x^*}{\partial \lambda}, \tag{35}$$

$$\frac{\partial x^*}{\partial \beta} = \frac{\Delta V(x^*)}{\beta}\frac{\partial x^*}{\partial \lambda}, \tag{36}$$

where $\Delta V(x)$ is given in eq. (30).

Let $0 < a < 1$ (negative social feedback). Then, $\frac{\partial x^*}{\partial \lambda} < 0$ where x^* may switch from stable to unstable and vice versa if λ increases.

- *If $x^* < \frac{1}{2}$ (i.e., $\Delta V(x^*) > 0$), $\frac{\partial x^*}{\partial \rho} > 0$ and $\frac{\partial x^*}{\partial \beta} < 0$.*
- *If $x^* > \frac{1}{2}$ (i.e., $\Delta V(x^*) < 0$), $\frac{\partial x^*}{\partial \rho} < 0$ and $\frac{\partial x^*}{\partial \beta} > 0$.*

For sufficiently large values of β and ρ, the equilibrium x^* becomes unstable.

Proof. See Appendix C. □

Proposition 4.1 is illustrated in Figure 11, which shows equilibria as a function of parameters (bifurcation diagram), with x^* if there is a unique stable equilibrium (single branch) and x_1^{**} and x_2^{**} if there is periodic dynamics (double branch).

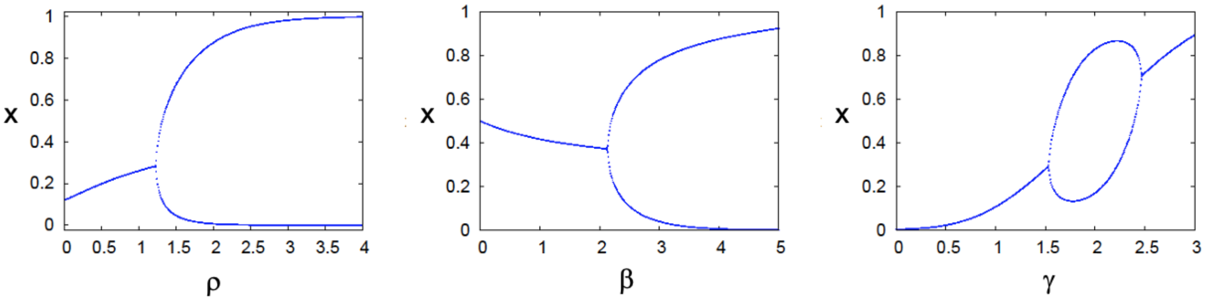

Figure 11: Long-run equilibria of x^* in the model with non-linear social interaction and negative feedback, assuming $a = 0.5$, $\delta_g = 2$, $\delta_b = 0$ and $\eta = 1$. Left: bifurcation diagram of ρ (for $\beta = 4$ and $\gamma = 1.5$). Centre: bifurcation diagram of β (for $\rho = 2$ and $\gamma = 1.5$). Right: bifurcation diagram of γ (for $\beta = 4$ and $\rho = 1.2$).

Consider first parameter β. If $\beta = 0$, $f(x)$ is a horizontal line with $f(x) = 1/2$ and so $x^* = 1/2$ (see Section 2). If β increases, $f(x)$ becomes steeper (with a negative slope) and x^* gradually decreases if $x^* < 1/2$ (which is the case in Figure 11, as $\Delta V(x^*) > 0$, given the parameter values assumed in this example), but would gradually increase if $\Delta V(x^*) < 0$ (not shown in Figure 11). As $f(x)$ becomes steeper (as does f^2) around the flex point and the equilibrium with increasing β, a value of β is reached where the equilibrium x^* becomes unstable. If β increases further, two stable equilibria x_1^{**} and x_2^{**} of f^2 emerge. (These equilibria behave exactly as the two equilibria of f in the basic model. See Proposition 3.3.) Any further increase of β moves x_1^{**} and x_2^{**} to the boundaries of $[0, 1]$. In other words, the amplitude of the oscillation increases.

For the degree of social interaction ρ, we also observe that an increase of ρ can cause a unique stable equilibrium to become unstable, with an increasing oscillation of the unstable equilibrium. However, and this is different from the basic model, if the equilibrium is stable, then with a negative feedback, equilibria are not reinforced, but just the opposite takes place. With a negative feedback, if we are in a brown equilibrium with $x^* < 1/2$ (as assumed in the example shown in Figure 11, left panel), with an increasing degree of social interaction ρ, the brown equilibrium becomes less attractive, as we are in a minority game. The opposite would be true if we were in a green equilibrium (not shown in Figure 11). Again, if ρ is sufficiently large, $f(x)$ becomes steep around the equilibrium (as does f^2), and the equilibrium x^* becomes unstable. This translates into f^2 having three equilibria, x^* being unstable and x_1^{**} and x_2^{**} being stable. As we know already for an upward sloping function $f(x)$, increasing social interaction moves x_1^{**} and x_2^{**} to the boundaries of the domain $[0, 1]$, and exactly this happens with f^2 too. Accordingly, the amplitude of the oscillation becomes larger, which shows up in the two branches in the left panel in Figure 11, which move to the boundaries of the domain of x as ρ increases.

Finally, a decrease of the composition parameter λ ($\lambda = \frac{\delta_g^2 - \delta_b^2}{2\eta} - \gamma$) increases the equilibrium fraction of individuals adopting a green lifestyle. In Figure 11, right panel, the warm glow parameter γ is gradually increased, which implies that λ is gradually decreased. The green lifestyle becomes increasingly attractive. Thus, x^* increases with γ. However, even though this parameter does not affect the slope of the function $f(x)$ (and also not of f^2), those functions change their location. This is shown in Figure 12 for $f(x)$ which travels from (a) to (f) if γ increases. For (a) and (b), as well as for (e) and (f), $f'(x) > -1$ and x^* is stable, whereas for (c) and (d), $f'(x) < -1$ and x^* is unstable. This shows up in Figure 11 where with increasing values of γ, x^* increases. The equilibrium travels first through a range of stability, then instability and then again stability.

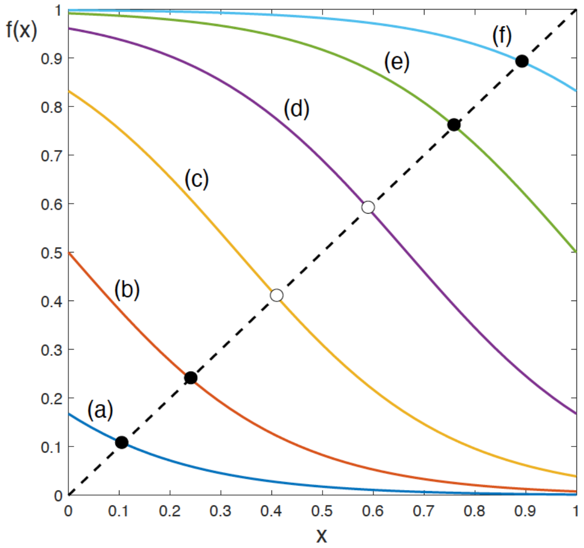

Figure 12: Examples of the map $f(x)$ for the model with non-linear social interaction and negative feedback for different values of the warm-glow parameter γ. (a) $\gamma = 1$, (b) $\gamma = 1.4$, (c) $\gamma = 1.8$, (d) $\gamma = 2.2$, (e) $\gamma = 2.6$, (f) $\gamma = 3$. The dots on the 45^o-line indicate stable (filled dot) and unstable (circled dot) equilibria. Other parameters are: $\beta = 4$, $\rho = 1.2$, $a = 0.5$, $\delta_g = 2$, $\delta_b = 0$, $\eta = 1$.

Thus, increasing warm glow of green consumers (i.e., γ increases) and/or the awareness of brown consumers (i.e., δ_b increases) implies a larger fraction of green consumers, also if there is negative feedback. However, what is now different is that there are no tipping points and there can be a transition phase with instability and periodic patters, before a stable equilibrium with more individuals following a green lifestyle is reached. Thus, with a negative feedback, transforming a society from brown to green is tricky undertaking.

Moreover, also with a negative feedback, increasing the environmental awareness of green consumers further is not helpful for the transition to a "green society". Campaigns to raise environmental awareness of the negative side effects of consumption should mainly address those individuals with little and not those with a high awareness.

5 Extension II: Endogenous Personal Norm

5.1 Endogenous Warm Glow

The utility which individuals derive from pursuing a green lifestyle was operationalized through a warm glow parameter γ in the basic model. Those adopting a green lifestyle receive an extra and constant benefit $\gamma > 0$. However, it is not unlikely that individuals measure their warm glow against the "average performance of society". If almost nobody adopts a green lifestyle, consuming environmentally responsible may give a higher personal satisfaction than if almost the entire population behaves “green”. In other words, the psychological well-being from pursuing a green lifestyle is not independent of social behaviour.

In our model, browns consume more than greens of an environmentally harmful good. We operationalise “relative warm glow” by letting the personal norm $G(x)$ no longer be a constant, but assume

$$G_i(x) = \begin{cases} \gamma(\langle q^* \rangle - q_g^*) & \text{if green,} \\ 0 & \text{if brown} \end{cases} \tag{37}$$

where γ is a (positive) parameter, which captures the intensity of warm glow, and where $\langle q^* \rangle = xq_g^* + (1-x)q_b^* = Q^*/n$ is the average consumption of society. Those who adopt a green lifestyle derive a warm glow which depends on relative consumption. The lower own consumption is compared to the society's average, the larger is the warm glow from the adoption of a green lifestyle. We note that Q^* is decreasing in x (see Proposition 2.1), and so does average consumption Q^*/n. Moreover, as the total quantity and the price move in the same direction in a market equilibrium (see again Proposition 2.1), p^* will decrease and individual consumption of greens will increase. Consequently, the well-being $G_i(x)$ from following a green lifestyle will decrease in x. Thus, this component in the payoff function is now associated with a negative feedback.

Assuming all components of the payoff function of individuals as in the basic model, except for the warm glow component $G_i(x)$, which is now given by eq. (37) instead of eq. (14), equilibrium individual consumption follows from the maximisation of payoff V with respect to quantity q_i in the second stage of the constituent game at time t, which leads to the first order conditions $U'(q_g^*) = p^*(x) + \delta_g + \gamma = 0$ for the green choice and $U'(q_b^*) = p^*(x) + \delta_b = 0$ for the brown choice in an interior equilibrium. This gives:

$$q_b^*(x) = \frac{v - \delta_b - p^*(x)}{\eta} \quad \text{and} \quad q_g^*(x) = \frac{v - \delta_g - p^*(x) - \gamma}{\eta}. \tag{38}$$

That is, the equilibrium demand of browns has not changed but that of greens is now lower compared to the basic model, which is evident by comparing eq. (38) with eq. (16).

As before, individuals are price takers and also "x-takers" in the second stage. However, as the warm glow depends on the choice of consumption level q_g, the first order condition of greens changes with respect to the basic model.[9] Accordingly, aggregate demand and the equilibrium price are different:

$$\begin{aligned} Q^*(p,x) &= n\Big[xq_g(x) + (1-x)q_b\Big] \\ &= \frac{n}{\eta}\left[v - p(x) - \delta_b - (\delta_g + \gamma - \delta_b)x\right]. \end{aligned} \tag{39}$$

$$p^*(x) = \frac{v - \delta_b - (\delta_g + \gamma - \delta_b)x}{1 + \eta\frac{s}{n}} \tag{40}$$

which is evident by consulting eqs. (17) and (19).

As stated in Section 2, the equilibrium fraction x is a fixed point of the probability distribution f of eq. (7), which depends on the discrete choice utility difference of browns and greens. Now, this difference is given by (see Appendix D.1):

$$\Delta V(x) \equiv V_b(x) - V_g(x) = \zeta' - \gamma(1-x)\frac{\delta_g + \gamma - \delta_b}{\eta} + \rho(1-2x), \tag{41}$$

where $\zeta' = \zeta + \frac{\gamma(2\delta_g+\gamma)}{2\eta}$ and $\zeta = \frac{\delta_g^2 - \delta_b^2}{2\eta}$. De facto, greens internalise consumption externalities not only with δ_g but now with $\delta_g + \gamma$. Thus, the difference in utility from material consumption of browns and greens is now larger because $\zeta' > \zeta$, and the warm glow for acting green is no longer a constant γ, but is $\gamma(1-x)\frac{\delta_g+\gamma-\delta_b}{\eta}$, which increases in γ but decreases in x.

It is clear from observing $\Delta V(x)$ in eq. (41) that an increase of warm glow γ has two opposite effects. On the one hand, it makes the brown lifestyle increasingly attractive, as greens receive a smaller utility from material consumption (i.e., $\Delta V(x)$ increases). On the other hand, it makes the green lifestyle more attractive due to a larger psychological well-being from following a green lifestyle for a given value of x (i.e., $\Delta V(x)$ decreases). Thus, the effect of the warm glow parameter γ on the equilibrium value of x is less straightforward than in the previous models (which is also due to the fact that the warm glow from a green lifestyle is a function of x itself). Also the effect of x on $\Delta V(x)$ is more complicated as we have $\Delta V'(x) = -2\rho + \gamma\frac{\delta_g+\gamma-\delta_b}{\eta}$ which can be either positive or negative.

Corollary 5.1. *In the model with endogenous warm glow (extension II), there is a critical level of the degree of social interaction*

$$\overline{\rho} = \gamma\frac{\delta_g + \gamma - \delta_b}{2\eta} \tag{42}$$

---

[9] Note that because the population is assumed to be large, the effect of q_g on average consumption is negligible and therefore ignored.

such that the lifestyle choice exhibits a positive feedback for $\rho > \overline{\rho}$ and a negative feedback for $\rho < \overline{\rho}$.

Depending on the relative strength of social utility and psychological well-being, $\Delta V(x)$ is increasing or decreasing in x. We obtain the first case if the social norm component is relatively strong, making the green lifestyle increasingly attractive with the number of individuals adopting a green lifestyle (positive feedback). The second case occurs if the psychological well-being component is relatively strong, rendering the green lifestyle less attractive if the number of individuals adopting a green lifestyle increases (negative feedback). Overall, the relative strength of these opposing effects is determined by the parameters ρ and γ. In Corollary 5.1, this is expressed in terms of the social interaction parameter ρ with the threshold $\overline{\rho}$, which is an increasing function of the personal well-being parameter from a green lifestyle γ.

The role of γ and ρ are even more evident from Corollary 5.2 and Figure 13.

Corollary 5.2. *A sufficient condition for a unique stable equilibrium is*

$$\overline{\rho} - \frac{2}{\beta} \leq \rho \leq \overline{\rho} + \frac{2}{\beta}. \tag{43}$$

A necessary condition for periodic dynamics is

$$\rho < \overline{\rho} - \frac{2}{\beta}. \tag{44}$$

A necessary condition for two equilibria is

$$\rho > \overline{\rho} + \frac{2}{\beta}. \tag{45}$$

Proof. See Appendix D.1. ☐

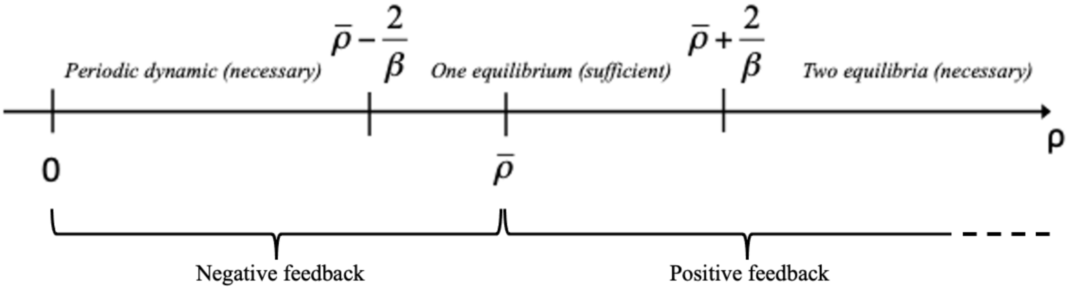

Figure 13: Different dynamic scenarios for the endogenous warm-glow model.

By gradually increasing the degree of social interaction ρ, the model "travels" along the ρ-axis from left to right in Figure 13. If ρ is sufficiently large, then by increasing gradually the warm glow parameter γ, we literally "travel" in the opposite direction. That is, the model may start in a regime with positive feedback and two equilibria and by increasing

γ, the threshold value $\overline{\rho}$ increases and therefore the model may belong to a new regime with only one equilibrium and a positive feedback. A further increase of γ places the model in the regime with a negative feedback.

In order to streamline the discussion, in what follows, we focus exclusively on the effects of the internalisation parameter of green and brown consumers, δ_g, δ_b, and the warm glow parameter γ on the equilibrium fraction of greens x^*. On this way, we first derive a result about the composition parameter ζ where we recall that $\zeta = \frac{\delta_g^2-\delta_b^2}{2\eta}$.

Lemma 5.1. *The effect of ζ on a stable equilibrium $x^* = f(x^*)$ is given by:*

$$\frac{\partial x^*}{\partial \zeta} = -\frac{\beta x^*(1-x^*)}{1+2\beta(\overline{\rho}-\rho)x^*(1-x^*)}. \tag{46}$$

The following cases in terms of the sign of (46) can be distinguished.

Part I: If $\rho < \overline{\rho}+\frac{2}{\beta}$, then $\frac{\partial x^*}{\partial \zeta} < 0$, regardless whether there is negative or positive feedback.

Part II: If $\rho > \overline{\rho}+\frac{2}{\beta}$ (positive feedback regime) and if there are two equilibria, then two characteristic values x_1^T and x_2^T exist, such that

(i) $\frac{\partial x^*}{\partial \zeta} < 0$ in the brown equilibrium $0 < x^* = x_b^* < x_1^T < \frac{1}{2}$ and in the green equilibrium $1 > x^* = x_g^* > x_2^T > \frac{1}{2}$;

(ii) Tipping points are attained at critical values ζ_1^T and ζ_2^T such that

  (a) $\lim_{\zeta\to\zeta_2^T} \frac{\partial x^*}{\partial \zeta} = -\infty$ with $\lim_{\zeta\to\zeta_2^T} x^* = x_2^T > \frac{1}{2}$;

  (b) $\lim_{\xi\to\xi_1^T} \frac{\partial x^*}{\partial \xi} = +\infty$ with $\xi = -\zeta$, $\xi_1^T = -\zeta_1^T$ and $\lim_{\xi\to\xi_1^T} x^* = x_1^T < \frac{1}{2}$.

Proof. See Appendix D.2 including the closed form solution of tipping points x_1^T and x_2^T. The existence of tipping points is assured by Theorem 1 in Section 2 □

Proposition 5.1. *The effect of δ_g on a stable equilibrium $x^* = f(x^*)$ is given by*

$$\begin{aligned}\frac{\partial x^*}{\partial \delta_g} &= \frac{\frac{\beta}{\eta}x^*(1-x^*)[\gamma(1-x^*)-\delta_g]}{1+2\beta(\overline{\rho}-\rho)x^*(1-x^*)}\\ &= \frac{\delta_g-\gamma(1-x^*)}{\eta}\frac{\partial x^*}{\partial \zeta}.\end{aligned} \tag{47}$$

A sufficient condition for $\frac{\partial x^*}{\partial \delta_g} < 0$ is $\delta_g > \gamma$, both in the unique equilibrium regime ($\rho < \overline{\rho}+\frac{2}{\beta}$) and in the two-equilibria regime ($\rho > \overline{\rho}+\frac{2}{\beta}$). In the two-equilibria regime, if the sufficient condition holds, at the tipping point x_2^T, a marginal increase of δ_g causes the green equilibrium to disappear.

The effect of δ_b on a stable equilibrium $x^* = f(x^*)$ is given by

$$\begin{aligned}\frac{\partial x^*}{\partial \delta_b} &= \frac{\frac{\beta}{\eta}x^*(1-x^*)(\delta_b-2\gamma+\gamma x^*)}{1+2\beta(\overline{\rho}-\rho)x^*(1-x^*)}\\ &= \frac{2\gamma-\gamma x^*-\delta_b}{\eta}\frac{\partial x^*}{\partial \zeta}.\end{aligned} \tag{48}$$

i) A sufficient condition for $\frac{\partial x^*}{\partial \delta_b} > 0$ *is* $\delta_b > 2\gamma$*, both in the unique equilibrium regime* ($\rho < \overline{\rho} + \frac{2}{\beta}$) *and in the two-equilibria regime* ($\rho > \overline{\rho} + \frac{2}{\beta}$)*. In the two-equilibria regime, if the sufficient condition holds, at the tipping point* x_1^T*, a marginal increase of* δ_b *causes the brown equilibrium to disappear.*

ii) A sufficient condition for $\frac{\partial x^*}{\partial \delta_b} < 0$ *is* $\delta_b < \gamma$*, both in the unique equilibrium regime* ($\rho < \overline{\rho} + \frac{2}{\beta}$) *and in the two-equilibria regime* ($\rho > \overline{\rho} + \frac{2}{\beta}$)*. In the two-equilibria regime, if the sufficient condition holds, at the tipping point* x_2^T*, a marginal increase of* δ_b *causes the green equilibrium to disappear.*

Proof. See Appendix D.2. □

The proposition confirms our conclusion from the basic model that raising the environmental awareness among the population may not always lead to the intended outcome. We found that a campaign turning greens into supergreens (i.e., δ_g increases) shifts the society to a browner equilibrium, even with the possibility (if there is a positive feedback and two stable equilibria) that a green tips over to a brown society at a social tipping point. In contrast, increasing the awareness of brown consumers about the externalities from consumption (i.e., δ_b increases) always lead to a greener society with the possibility that a brown society suddenly jumps to a green society at a social tipping point.

Now, in extension II, a new scenario arises in result (ii). Only if the environmental awareness of the population which has adopted a brown lifestyle is already sufficiently large ($\delta_b > 2\gamma$), will a further increase of awareness have the desired effect, namely to increase the population which adopts a green lifestyle and (if there is positive feedback) eventually tipping the society to a green equilibrium. However, if the awareness about the negative environmental effects of consumption of individuals following a brown lifestyle is very low, increasing their awareness does not have the intended effect, which appears paradoxically.

Proposition 5.2. *The effect of* γ *on a stable equilibrium* $x^* = f(x^*)$ *is given by*

$$\frac{\partial x^*}{\partial \gamma} = -\frac{\gamma(1-2x) - (\delta_g - \delta_b)x - \delta_b}{\eta} \frac{\partial x^*}{\partial \zeta}. \tag{49}$$

In the unique equilibrium regime ($\rho < \overline{\rho} + \frac{2}{\beta}$) *and in the two-equilibria regime with positive feedback* ($\rho > \overline{\rho} + \frac{2}{\beta}$)*, we have* $\frac{\partial x^*}{\partial \gamma} < 0$ *in a green equilibrium* $x_g^* > \frac{1}{2}$*.*

In the two-equilibria regime, at the tipping point x_2^T*, with a corresponding critical value* γ^T*, we have*

$$\lim_{\gamma \to \gamma^T} \frac{\partial x^*}{\partial \gamma} = -\infty, \quad \textit{with} \quad \lim_{\gamma \to \gamma^T} x^* = x^T > \frac{1}{2}.$$

Proof. See Appendix D.2. □

According to Proposition 5.2, and as already indicated above, the effect of an increasing warm glow parameter γ is not always clear-cut for extension II, as this was true in the basic model. In the basic model, increasing warm glow always lead to larger fraction of the population which adopted a green lifestyle, including the possibility to tip a brown to a green society. Now, in extension II, this is also possible in a brown equilibrium.[10] However, in an already green society, increasing warm glow has just the opposite and undesirable effect. The brown lifestyle becomes more attractive and a green equilibrium may even tip to a brown equilibrium.

In Figure 14, this interesting case where an increase of warm glow in a green equilibrium reduces the equilibrium number of individuals following a green lifestyle is illustrated. We start in scenario (a), assuming the inhabited equilibrium is green, $x_g^* \simeq 1$. Increasing γ from 0 up to the bifurcation value $\gamma^T = 0.68$, we arrive at scenario (b), where the map $f(x)$ is tangent to the 45^o line, and the equilibrium attains a tipping point. Increasing γ further makes the population shift behaviour to the only remaining stable equilibrium, which is a brown equilibrium with $x_b^* \simeq 0$ in scenario (c). This implies a drastic drop of greens in the population. For $\gamma = 4$, in scenario (d), the system is in a regime with negative feedback, with a unique stable brown equilibrium. Finally, in scenario (e), with $\gamma = 5$, the equilibrium is unstable and choices in the population are characterised by periodic dynamics. As Proposition 5.2 refers to stable equilibria, scenario (e) is not captured.

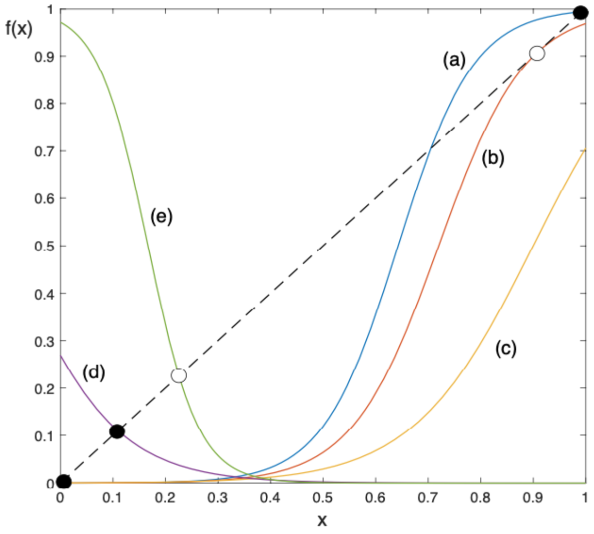

Figure 14: Examples of the map $f(x)$ for the model with endogenous warm-glow with different levels of intensity of the personal norm γ. (a) $\gamma = 0$, (b) $\gamma = 0.68$, (c) $\gamma = 1.5$, (d) $\gamma = 4$, (e) $\gamma = 5$. The dots on the 45^o-line indicate stable (filled dot) and unstable (circled dot) equilibria. Case (b) is a tangent bifurcation that represents a social tipping point. Other parameters are $\beta = 1$, $\rho = 7$, $\delta_g = 2$, $\delta_b = 0$, $\eta = 1$.

---

[10] As we cannot establish a sufficient condition for this case to arise, it is not mentioned in Proposition 5.2. However, in a brown equilibrium for x^* sufficiently close to zero, $\frac{\partial x^*}{\partial \gamma} > 0$ is possible, as this is evident when moving from (c) to (d) in Figure 14.

6 Consumption Patterns, Global Damages and Social Welfare

6.1 Preliminaries

In this section, we analyse how important parameters of our model influence individual and global consumption patterns, global damages and global welfare. Global consumption is directly linked to global damages, which is just the sum over all n individual damages where individual damages are given by $D(Q) = \delta Q$ (see eq. (13)) in all model versions. We notice that total consumption $Q = n(xq_g(x) + (1-x)q_b(x))$ as well as individual consumption $q_g(x)$ and $q_b(x)$ are a function of the parameters of the model and x. This suggests to group parameters into three categories.

In the first category are parameters which only have a direct effect on individual and total consumption like s, n and v in all model versions.

In the second category are parameters which only have an indirect effect on consumption via a change of x. For instance, in the basic model and extension I, ρ, β and γ belong to this group. In extension 2, only the first two parameters are in this category.

In the third category are parameters which have a direct and indirect effect on consumption, like parameters η, δ_g and δ_b in all model versions. For extension II, also the warm glow parameter γ belongs to this category.

It appears sensible to measure social welfare in pure economic terms. That is, we exclude the social utility I_i and warm glow component G_i in individual payoffs:

$$SW(x) = \sum_{i=1}^{n} CS_i(q_i^*(x)) + PS(Q^*(x)) - \sum_{i=1}^{n} D_i(Q^*(x)),$$

where CS_i stands for consumer surplus, which is aggregate utility over green and brown consumers minus expenses, PS stands for the aggregate producer surplus, which is total revenue minus production costs, and the final term are aggregate or total damages which we henceforth abbreviate $TD = \sum_{i=1}^{n} D_i(Q^*(x))$. All welfare components are a function of x and the parameters of the model. Environmental damages are evaluated with the "objective" damage parameter δ.

In Appendix E, we show that for all three model versions which we presented in sections 3, 4 and 5, social welfare $SW(x)$ increases in x, provided environmental damages receive a sufficiently high weight in the social welfare function, i.e., $\delta > \underline{\delta}$ in the basic model and extension I and $\delta > \underline{\underline{\delta}}$ in extension II. This makes perfectly sense. On the one hand, consumer surplus and producer surplus are equal to total utility derived from consumption minus production cost. This aggregate "net utility" increases in consumption levels. Since brown consume more than green consumers, this component is negatively affected by the number of individuals choosing a green lifestyle. On the other hand, environmental damages are reduced by lower aggregate consumption levels. Subsequently, we assume that the sufficient conditions for $\frac{\partial SW(x)}{\partial x} > 0$ hold where $x^* = 1$ maximises social welfare.

Subsequently, we analyze how key parameters affect consumption, environmental damges and social welfare. We start with the basic model in subsection 6.2 and subsequently evaluate what changes in Extensions I and II in subsection 6.3.

6.2 Basic Model

The first category of parameters is less interesting, as effects are straightforward. As mentioned, all these parameters do not affect x^* (see eq. (21)). The larger s, the lower are marginal production costs (see eq. (19)), the lower will be the equilibrium market price (see eq. (19)), and, therefore, the larger will be the total quantity consumed for any value of x^* (see eq. (17)). This is also evident from the market equilibrium condition $Q(x) = sp(x)$. Similarly, the larger the population size n and the larger the intercept of the marginal utility at zero consumption v and the larger the total quantity consumed for any given x^* (see eq. (17)). We recall that higher total consumption implies higher global environmental damage in our model. It is not sensible to make interference with respect to social welfare as all parameters in this category are parameters in the social welfare function itself.

For the second category of parameters, a more interesting pattern is observed. Conclusions about social welfare are possible as parameters in this category only affect x^*.

Proposition 6.1. *An increase of the warm glow parameter γ, the intensity of social interactions ρ and the intensity of choices β implies the following:*

- $\frac{\partial Q^*}{\partial \gamma} < 0$, $\frac{\partial q_i^*}{\partial \gamma} > 0$, $\frac{\partial TD^*}{\partial \gamma} < 0$ *and* $\frac{\partial SW^*}{\partial \gamma} > 0$ *with* $i = b, g$, *regardless whether the population is in a stable green or brown equilibrium. At a tipping point, the brown equilibrium disappears, such that total consumption and environmental damages decrease and social welfare increases abruptly.*

- *Let* $\phi = \beta, \rho$*:*

  $\frac{\partial Q^*}{\partial \phi} > 0$, $\frac{\partial q_i^*}{\partial \phi} < 0$, $\frac{\partial TD^*}{\partial \phi} > 0$ *and* $\frac{\partial SW^*}{\partial \phi} < 0$ *in a stable brown equilibrium* $(x_1^* < 1/2)$*,*

  $\frac{\partial Q^*}{\partial \phi} < 0$, $\frac{\partial q_i^*}{\partial \phi} > 0$, $\frac{\partial TD^*}{\partial \phi} < 0$ *and* $\frac{\partial SW^*}{\partial \phi} > 0$ *in a stable green equilibrium* $(x_2^* > 1/2)$*.*

Proof. See Appendix F, including the consideration of social tipping points. □

An increase in the warm glow parameter γ makes the green lifestyle more appealing in the basic model, which translates into an increase of x^* (Proposition 3.2). If x^* increases, social welfare increases, aggregate consumption Q^* and global damages TD^* decrease and so does the market price p^* (see eqs. (17) and (19)). Consequently, individual consumption levels increase, both for greens and browns (see eq. 16). The fact that total consumption decreases despite individual consumption increases is because the composition effect (more greens than browns in the population with greens consuming less than browns) is stronger than the behavioural effect (the increase in individual consumption levels). For the basic model, we also know that if γ increases sufficiently enough, a threshold value may be reached where the brown equilibrium may disappear (see Proposition 3.2). This certainly decreases total consumption, global environmental damages and increases social welfare in a non-continuous way.

The same mechanism is at play for a change in social interaction parameter ρ or a change in the intensity of choices β (see Proposition 3.3). An increase of both parameters "reinforces" choices and does not allow for social tipping points. (They would arise if these parameters were lowered.) Hence, the direction of change in consumption levels depends on whether the population is in a green equilibrium ($x^* > 1/2$) or in a brown equilibrium ($x^* < 1/2$). In the first case, aggregate consumption decreases while individual consumption increases, whereas in the second case this is reversed. In both cases, the effect of parameters on the change of fractions of greens and browns in the population (composition effect) is stronger than the change of individual behaviour. The change of aggregate consumption goes in the opposite direction of social welfare, given our assumption that environmental damages receive a sufficiently high weight in the social welfare function.

Taken together and interestingly, even though the second class of parameters have no direct effect on consumption levels, but only an indirect effect through a change of the fraction of greens and browns in the population, the resulting effect does not only change the composition of the population, but has also an effect on individual behaviours through a price change.

We are now turning to the third category of parameters, which have a direct and indirect effect and thus effects are more intricate. We only discuss the effect of an increase of the internalisation parameter of browns and greens, δ_b and δ_g, respectively. The parameter η is not considered as it is part of the social welfare function itself.

Proposition 6.2. *An increase of the internalisation parameter of brown choices* δ_b*, both in a stable brown and green equilibrium, causes the number of greens to increase,* $\frac{\partial x^*}{\partial \delta_b} > 0$*, aggregate consumption and global environmental damages to decrease, i.e.,* $\frac{\partial Q^*}{\partial \delta_b} < 0$ *and* $\frac{\partial TD^*}{\partial \delta_b} < 0$*, green consumption to increase, i.e.,* $\frac{\partial q_g^*}{\partial \delta_b} > 0$*, social welfare to increase, i.e.,* $\frac{\partial SW^*}{\partial \delta_b} < 0$ *(if* $\delta > \frac{\delta_b}{n}$*) and the following scenarios for brown consumption to occur:*

(I) $-1 < \frac{\partial p^*}{\partial \delta_b} < 0$ *(weak price effect), then* $\frac{\partial q_b^*}{\partial \delta_b} < 0$*,*

(II) $\frac{\partial p^*}{\partial \delta_b} < -1$ *(strong price effect), then* $\frac{\partial q_b^*}{\partial \delta_b} > 0$*.*

An increase of δ_b *at a social tipping point will make the brown equilibrium to disappear such that global consumption and environmental damages drop abruptly and social welfare increases suddenly.*

Proof. See Appendix F, including the consideration on social tipping points. □

An increase of δ_b reduces the material utility from consumption of browns and thus renders the brown lifestyle less attractive (see eq. (21)). Thus, more individuals choose a green lifestyle (see Proposition 3.2), which reduces total consumption (see eq. (17)) and hence global environmental damages, despite greens consume now more, as the equilibrium price has dropped (see eq. (16)). This may be seen as an unintentional side effect which is transmitted through the market. If the price effect is strong, even browns may increase individual consumption. However, the composition effect of the population always dominates the individual behaviour effect such that total consumption decrease and hence environmental damages as well. Thus, raising the awareness of brown consumers about the externality, which they cause, reduces global environmental damages and increases social welfare.

Proposition 6.3. *An increase of the internalisation parameter of green choices* δ_g*, both in a stable brown and green equilibrium, causes the number of greens to decrease, i.e.,* $\frac{\partial x^*}{\partial \delta_g} < 0$ *with the following scenarios to occur:*

(I) If $\frac{x^*}{\delta_g - \delta_b} > \left|\frac{\partial x^*}{\partial \delta_g}\right|$*, then* $\frac{\partial p^*}{\partial \delta_g} < 0$*,* $\frac{\partial Q^*}{\partial \delta_g} < 0$*,* $\frac{\partial TD^*}{\partial \delta_g} < 0$*,* $\frac{\partial q_b^*}{\partial \delta_g} > 0$ *and* $\frac{\partial q_g^*}{\partial \delta_g} >, < 0$*.*

(II) If $\frac{x^*}{\delta_g - \delta_b} < \left|\frac{\partial x^*}{\partial \delta_g}\right|$*, then* $\frac{\partial p^*}{\partial \delta_g} > 0$*,* $\frac{\partial Q^*}{\partial \delta_g} > 0$*,* $\frac{\partial TD^*}{\partial \delta_g} > 0$*,* $\frac{\partial q_b^*}{\partial \delta_g} < 0$ *and* $\frac{\partial q_g^*}{\partial \delta_g} < 0$*.*

An increase of δ_g *at a social tipping point will make the green equilibrium to disappear and will increase total consumption and environmental damages abruptly. Social welfare may decrease with an increase of* δ_g *and will so most likely at a social tipping point.*

Proof. See Appendix F, including the consideration on social tipping points. □

If the awareness of greens about the externality caused by consumption increases, they reduce consumption. This is the direct effect on consumption which is the left-hand side term in condition (I) and (II) stated above. The utility from material consumption of greens decreases which makes the brown lifestyle more attractive. This is the indirect effect on the composition of the population which is the right-hand side term in condition (I) and (II) above. Hence, the fraction of greens in the population decreases. If condition (I) holds, the direct effect is stronger than the indirect effect and total consumption will decrease, even though browns will consume more due to a lower price. However, if condition (II) holds, then increasing environmental awareness has the opposite and intriguing effect that total consumption and thus environmental damages increase. The indirect composition effect is so strong, i.e., the number of browns in the population increases so much, and due to the fact they consume more than greens, total consumption increases. This is also what happens at a social tipping point: if the population is in a green equilibrium, this equilibrium will disappear with a marginal increase of δ_g. This abrupt change of x^* has the effect of case (II) above, as $\frac{\partial x^*}{\partial \delta_g} = -\infty$.

Below, we consider two examples which illustrate Proposition 6.3. Example 1 belongs to case II and assumes the following parameter values: $\delta_b = 0.1$, $\delta_g = 0.4$, $\eta = 1$, $\gamma = 0.025$ such that $\zeta = 0.05$; $\beta = 10$ and $\rho = 0.1$ such that, due to $\rho\beta = 1$, we have a unique brown equilibrium. For $n = 1000$, $s = 500$ and $v = 2$, we have: $x^* = f(x^*) \simeq 0.2815$. Moreover, $p^* = 1.2103$, $q_b^* = 0.6897$, $q_g^* = 0.3897$ and $Q^* = 605.15$. Social welfare is given by $SW^* = 651.89 - 605183.33\delta$ where we recall that δ is the damage parameter that evaluates global damages.

If the internalisation parameter increases to $\delta_g = 0.6$, then $\zeta = 0.15$. We find for the equilibrium values: $x^* = f(x^*) \simeq 0.0894$, $p^* = 1.2369$, $q_b^* = 0.6631$, $q_g^* = 0.1631$ and $Q^* = 618.45$. Even though individual consumption has decreased (individual behaviour), aggregate consumption has increased, as the fraction of greens has dropped substantially (composition effect). Thus, aggregate environmental damages will be larger. Social welfare is given by $SW^* = 653 - 618433.33\delta$. Thus, if the damage parameter δ is sufficiently large, social welfare will have decreased. As we assume $\delta > \underline{\delta}$ (see Section 6.1), this is indeed the case.

The second example belongs initially to case I, but with a further increase of δ_g, it subsequently captures the result about tipping points. Example 2 assumes initially the following parameters values: $\delta_b = 0.1$, $\delta_g = 0.4$, $\eta = 1$ and $\gamma = 0.025$ such that $\zeta = 0.05$; $\beta = 10$, $\rho = 0.5$ such that $\rho\beta = 5$, which gives two stable equilibria, which for $n = 1000$, $s = 500$ and $v = 2$ are given by $x_1^* = f(x_1^*) \simeq 0.0042$ and $x_2^* = f(x_2^*) \simeq 0.9876$.

Assume the second green equilibrium is populated. Then the equilibrium values are given by: $p(x_2^*) = 1.0692$, $q_b(x_2^*) = 0.8308$, $q_g(x_2^*) = 0.5308$, $Q(x_2^*) = 534.6$ and $SW(x_2^*) = 639.94 - 534573.33\delta$.

By increasing the internalisation parameter to $\delta_g = 0.6$, we have $\zeta = 0.15$ and the green equilibrium values are given by: $x_2^* = f(x_2^*) \simeq 0.9546$, $p(x_2^*) = 0.9485$, $q_b(x_2^*) = 0.9515$, $q_g(x_2^*) = 0.4515$ $Q(x_2^*) = 474.25$ and $SW(x_2^*) = 605.7 - 474233.33\delta$. This means that aggregate consumption and hence environmental damages have decreased, which is case I in Proposition 6.3. Moreover, social welfare has increased as we assume the condition

$\delta > \underline{\delta}$ to hold.

However, a further increase of the internalisation parameter of greens to $\delta_g = 0.7$, implying $\lambda = 0.215$, implies that the system passes a social tipping point. That is, the population 'jumps' to the following (unique) brown equilibrium: $x^* = f(x^*) \simeq 0.0008$, $p^* = 1.2667$, $q_b^* = 0.6333$, $q_g^* = 0.0333$, $Q^* = 633.35$ and $SW^* = 664.84 - 633173.33\delta$. After the tipping point, aggregate consumption has increased to a level which is even higher than in the initial situation. Hence, environmental damages are also higher and social welfare lower than in the initial situation, given that we assume $\delta > \underline{\delta}$.

Obviously, it would be easy to construct cases where not only the internalisation parameter of greens but also of browns increase at the same time, even though that of greens increases more pronounced, with the same effects as just observed. Thus, a policy aimed at increasing the perception of environmental damages among its citizens may have not the intended effect. This is not an unrealistic scenario: environmental awareness campaigns are often characterised by a 'self-selection' mechanism: people who already behave environmentally friendly tend to more receptive to this campaigns than others. Such a self-selection mechanism may translate into higher consumption and environmental damages and lower social welfare despite environmental awareness of the population has increased.

6.3 Extensions I and II

For extensions I and II, the effect of changes in parameters of the first category with only a direct effect would be exactly the same. For extension I, with non-linear social norms, all Propositions (namely Propositions 6.1, 6.2, and 6.3) hold if there is a positive decision feedback. If there is a negative feedback, which was the focus in Section 4, then this is also true, with three qualifications. First, we need to assume a unique stable equilibrium (no periodic dynamics). Second, there are no tipping points. Third, the effect of the social interaction parameter ρ on total and individual consumption has just the opposite sign than in Proposition 6.1. An increase of ρ does not have a reinforcing effect on x^* (which decreases in a brown and increases in a green equilibrium) but has just the opposite effect on x^*, Q^*, TD^* and SW^*.

For extension II, with endogenous personal norms, it is no longer true that the warm glow parameter γ has only an indirect effect and hence belongs to the second group of parameters. It also has now a direct effect on consumption and so belongs to the third group of parameters, which is evident from $q_g^*(x)$ given in eq. (38). Thus, Proposition 6.1 would be different. Increasing warm glow does not always lead to a larger fraction of greens in the population with lower total consumption and the possibility of a brown equilibrium tipping to a green equilibrium. From Proposition 5.2 we know that an increase of warm glow may well have the opposite effect: the fraction of the population adopting a brown lifestyle may increase and a green may tip to a brown equilibrium. Consequently, one can easily perceive that total consumption and hence environmental damages may well increase. Even though greens may consume less, browns consume more and there are now more people with a brown consumption behaviour. Thus, social welfare may well decrease.

This is very similar phenomenon as described for the basic model regarding the internalisation parameter of green consumers δ_g. According to Proposition 6.3, a higher environmental awareness of greens can lead to more total consumption and a marginal increase of δ_g may cause the green equilibrium to disappear at a tipping point, with an abrupt increase of total consumption and global damages and a drop of social welfare. This surprising result would also hold for extension II regarding the warm glow parameter γ.

Finally, in terms of the internalisation parameter of brown consumers δ_b, the clear-cut results of Proposition 6.2 would be more nuanced for extension II. This is because an increase of δ_b may not only result into a larger equilibrium fraction of green consumers, but also into a smaller fraction and may not only lead to the disappearance of a brown but also of a green equilibrium according to Proposition 5.2. That is, an increase of δ_b does not always lead to a reduction of total consumption and environmental damages with a possible abrupt change at a brown tipping point. Also, the completely reverse result is possible.

From the discussion, it is evident that for extension II some new aspects with none obvious results emerge. The effect of an increase of warm glow of green consumers and an increase of environmental awareness of brown consumers on the split between green and brown consumers, total consumption and global damages as well as social welfare crucially depends on the starting conditions and on the relative values of these and other parameters of the model. General concusions are far more difficult than for the base model and extension I.

7 Conclusion

Our model brings discrete choice and social interactions into environmental economics, and focuses on the demand side to characterise the factors that determine a prevalence of sustainable or unsustainable consumption behaviour. In particular, the model allows to describe quantitatively factors such as social norms, psychological well-being (warm-glow) and environmental damage preference.

The evolutionary approach of a large population of decision makers enables us to aggregate individual decisions and to obtain different scenarios as emergent properties. Moreover, we can study the model both in terms of its equilibria, and also as a dynamic system, and we can make a stability analysis in different conditions.

The intensity of social interactions is the main factor of our model. We obtain a number of analytical results that characterise four different regions in the social interactions space. Threshold values of social interactions intensity depend on other factors such as the size of psychological well-being, the elasticity of individual demand, and the marginal environmental damage. The mathematical characterisation of these factors suggest a number of messages for policy action, where the different parameters are viewed as different channels for governmental intervention.

The contribution of this paper to environmental economics is twofold: first, it offers a novel approach that focuses on the demand side, studying the determinants of consumers behaviour and analysing quantitatively the interplay of individual decisions, social norms and the environment. Second, the model offers a quantitative description of factors that inhabit the social psychology literature, and in particular environmental psychology.

A number of extensions can be envisaged for our model. First, different factors in the decision problem can be complicated, in order to account for crowding-out effects, and for dynamic psychological well-being. We can also focus on the expectations formation process of the behaviours fraction, and study heterogeneous forecasting rules. Finally the model can be framed in a overlapping generation framework, and the decision system may be extended to environmental preferences across generations.

We believe that a discrete choice framework, the population approach, and the dynamical system version of the model can be a useful tool to study and understand the conditions for a transition to sustainable lifestyles.

Appendix A Details of the Proof of Proposition 2.2

Existence of equilibrium. The map $f(x)$ is a monotonic transformation of a linear function $\Delta V(x)$, resulting in a monotonic and continuous function $f : [0,1] \longrightarrow [0,1]$. By applying the intermediate value theorem, there must exist one value $x^* = f(x^*) \in [0,1]$.

Uniqueness of equilibrium. Here we distinguish the cases of f increasing or decreasing in x, starting with f increasing.

If $f'(x^F) \leq 1$, we automatically have uniqueness since x^F is the point where f obtains its largest slope. For any fixed point x^* of f, $f'(x^*) < 1$ must hold. Computing the first derivative of f in (8) and evaluating it at x^F, we derive $|\Delta V'(x)| \leq \frac{4}{\beta}$ in eq. (10). If $x^F \notin [0,1]$, uniqueness of an equilibrium follows: $f(x)$ is a logistic S-shaped function. Hence, there is only one flex point x^F, such that $f''(x^F) = 0$ holds. This is where $f'(x)$

obtains its maximum value. If $f'(x^F) \leq 1$, uniqueness is obvious. If $f'(x^F) > 1$, we can find a point $\overline{x}$ such that $f'(\overline{x}) = 1$. Let us consider $x^F < 0$. Then two cases are possible, either $\overline{x} < 0$ or $\overline{x} > 0$. If $\overline{x} < 0$, and because $f(x)$ is concave for $x > x^F$, we have $f'(x) < 1$ in $[0, 1]$, and, hence, uniqueness. If $\overline{x} > 0$, we have $f'(x) > 1$ in $[0, \overline{x}]$. But we have $f(x) > x$ in $[0, \overline{x}]$ because $f'(x) > 1$ and $f(x) > 0$, as $x^F < 0$ by assumption. This rules out $f(x)$ crossing the 45^o-line with slope $f'(x) > 1$ in [0,1]. It can only cross it with $f'(x) < 1$ in $[\overline{x}, 1]$, and only once because of the concavity for $x > x^F$, which means that we can only have one stable equilibrium. By the same token, we can address the case $x^F > 1$. Due to symmetry of f with respect to the flex point, $f(x - x^F) = 1 - f(x^F - x)$, this is straightforward. Notice that $f'(x^F) \leq 1$ is a sufficient condition for uniqueness of a stable equilibrium because if $x^F = x^*$ is the fixed point, it is stable; if x^F is not the fixed point, then $|f'(x)| < 1 \ \forall x \neq x^F$.

Multiplicity of equilibria. If there is a fixed point $x^* \in [0, 1]$, such that $f'(x^*) > 1$, there must be two stable equilibria, i.e., $x_b^* < 1/2$ and $x_g^* > 1/2$. First, we notice that $f(x)$ is continuous; second, $f(x) \in [0, 1]$ and, third, f is S-shaped. The last feature implies that f crosses the 45^o-line in x^* from below. Hence, it needs to cross it also in two other points, namely, $x_b^* < 1/2$ and $x_g^* > 1/2$, with $f'(x_b^*) < 1$ and $f'(x_g^*) < 1$. Finally, $\Delta V(x^F) = 0 \Rightarrow f(x^F) = 1/2$, implying $x_b^* < 1/2$ and $x_g^* > 1/2$ due to the monotonicity of f.

If f is a decreasing function over the entire domain of x, uniqueness is obvious, as only one fixed point can exist. A sufficient condition for uniqueness is $|\Delta V'(x)| \leq \frac{4}{\beta}$. Notice that $x^F \notin [0, 1]$ is not sufficient for a stable equilibrium if f is decreasing in x.

Appendix B Proofs of the Basic Model

B.1 The Effect of λ on x^*

Starting from the fixed point condition of the logistic map in eq. (7), we compute the derivative with respect to λ on both sides:

$$\frac{\partial x^*}{\partial \lambda} = \frac{\partial}{\partial \lambda} \frac{1}{1 + e^{\beta \Delta V(x^*)}}$$

where ΔV is given by eq. (21). For notational simplicity, we use x for x^*, and ΔV for $\Delta V(x)$. Let $x_\lambda = \frac{\partial x^*}{\partial \lambda}$, then we have

$$x_\lambda = -\frac{\beta \frac{\partial \Delta V}{\partial \lambda} e^{\beta \Delta V}}{\left(1 + e^{\beta \Delta V}\right)^2} = \frac{\beta e^{\beta \Delta V} \left(2\rho x_\lambda - 1\right)}{\left(1 + e^{\beta \Delta V}\right)^2}.$$

Rearranging, we get

$$x_\lambda = \frac{\beta e^{\beta \Delta V}}{2\rho \beta e^{\beta \Delta V} - \left(1 + e^{\beta \Delta V}\right)^2}.$$

This can be rewritten in terms of $x = \frac{1}{1+e^{\beta\Delta V(x)}}$ as:

$$x_\lambda = \frac{\beta\frac{1-x}{x}}{2\beta\rho\frac{1-x}{x} - \frac{1}{x^2}} = \frac{\beta(1-x)x}{2\rho\beta(1-x)x - 1}. \tag{50}$$

This expression becomes arbitrarily large when the denominator approaches zero. This occurs when x approaches the solutions of $2\rho\beta(1-x)x = 1$. This gives us two bifurcation points to which we refer as social tipping points.

$$x_1^T = \frac{1 - \sqrt{1 - \frac{2}{\rho\beta}}}{2}, \quad x_2^T = \frac{1 + \sqrt{1 - \frac{2}{\rho\beta}}}{2}. \tag{51}$$

From expression (50), we see that $x_\lambda < 0 \Leftrightarrow 2\rho\beta(1-x)x < 1$. Since $x(1-x)$ is a reversed parabola, we have $2\rho\beta(1-x)x < 1$ and $x_\lambda < 0$ for $x < x_1^T$ and $x > x_2^T$. For $x_1^T < x < x_2^T$, we have $2\rho\beta(1-x)x > 1$ and $x_\lambda > 0$.

We can compute the first derivative of the map f when the equilibrium is in a tangent bifurcation. For $x^* = x_1^T$, we have

$$\begin{aligned} f'(x^* = x_1^T) &= 2\rho\beta\left(1 - \frac{1 - \sqrt{1 - \frac{2}{\rho\beta}}}{2}\right)\frac{1 - \sqrt{1 - \frac{2}{\rho\beta}}}{2} \\ &= \frac{\rho\beta}{2}\left(1 + \sqrt{1 - \frac{2}{\rho\beta}}\right)\left(1 - \sqrt{1 - \frac{2}{\rho\beta}}\right) = \frac{\rho\beta}{2}\frac{2}{\rho\beta} = 1. \end{aligned} \tag{52}$$

Similarly, we can show that $f'(x^* = x_2^T) = 1$. In order to understand the type of bifurcation, we need to look at the second derivative of the map f, given by eq. (8).

Generally, for $\Delta V(x^*) > 0$, $x^* < x^F$ and $f''(x^*) > 0$. That is, $f(x)$ is a convex map at the tangency point. For $\Delta V(x^*) < 0$, we have $f''(x^*) < 0$. That is, $f(x)$ is a concave map at the tangency point. In both cases, we have a bifurcation in x^* where the map f is tangent to the 45^o-line. This is the condition that characterises a tangent bifurcation.

Bifurcation values of the parameter λ are obtained by requiring that a tipping point x_1^T or x_2^T is a fixed point:

$$\begin{aligned} x_1^T &= \frac{1 - \sqrt{1 - \frac{2}{\rho\beta}}}{2} = \frac{1}{1 + e^{\beta(\lambda + \rho(1 - 2x_1^T))}}, \\ x_2^T &= \frac{1 + \sqrt{1 - \frac{2}{\rho\beta}}}{2} = \frac{1}{1 + e^{\beta(\lambda + \rho(1 - 2x_2^T))}}. \end{aligned} \tag{53}$$

We compute bifurcation values λ_1^T and λ_2^T. Rearranging gives:

$$e^{\beta(\lambda_1^T + \rho(1-2x_1^T))} = \frac{1 - x_1^T}{x_1^T} \quad \Rightarrow \quad \lambda_1^T + \rho(1 - 2x_1^T) = \frac{1}{\beta}\ln\left(\frac{1 - x_1^T}{x_1^T}\right)$$

$$e^{\beta(\lambda_2^T + \rho(1-2x_2^T))} = \frac{1 - x_2^T}{x_1^T} \quad \Rightarrow \quad \lambda_2^T + \rho(1 - 2x_2^T) = \frac{1}{\beta}\ln\left(\frac{1 - x_2^T}{x_2^T}\right)$$

and by substituting the expressions of x_1^T and x_2^T from eq. (53), we have

$$\lambda_1^T = \frac{1}{\beta}\ln\left(\frac{1+\sqrt{1-\frac{2}{\rho\beta}}}{1-\sqrt{1-\frac{2}{\rho\beta}}}\right) - \rho\sqrt{1-\frac{2}{\rho\beta}} \tag{54}$$

$$\lambda_2^T = \frac{1}{\beta}\ln\left(\frac{1-\sqrt{1-\frac{2}{\rho\beta}}}{1+\sqrt{1-\frac{2}{\rho\beta}}}\right) + \rho\sqrt{1-\frac{2}{\rho\beta}}. \tag{55}$$

Notice that

$$\begin{aligned}\lambda_2^T &= \frac{1}{\beta}\ln\left(\frac{1+\sqrt{1-\frac{2}{\rho\beta}}}{1-\sqrt{1-\frac{2}{\rho\beta}}}\right)^{-1} + \rho\sqrt{1-\frac{2}{\rho\beta}} \\ &= -\frac{1}{\beta}\ln\left(\frac{1+\sqrt{1-\frac{2}{\rho\beta}}}{1-\sqrt{1-\frac{2}{\rho\beta}}}\right) + \rho\sqrt{1-\frac{2}{\rho\beta}} = -\lambda_1^T. \end{aligned} \tag{56}$$

B.2 The Effect of β on x^*

We compute the derivative with respect to β on both sides of the fixed point condition for the logistic map in eq. (7). Let $x_\beta = \frac{\partial x^*}{\partial \beta}$, then we have:

$$x_\beta = -\frac{1}{\left(1+e^{\beta\Delta V}\right)^2}\left[e^{\beta\Delta V}\left(\Delta V + \beta\frac{\partial \Delta V}{\partial \beta}\right)\right].$$

Defining $A = \left(1+e^{\beta\Delta V}\right)$ and $E = e^{\beta\Delta V}$, as $\frac{\partial \Delta V}{\partial \beta} = -2\rho x_\beta$, the equation above can be written as

$$x_\beta = -\frac{E}{A^2}\left(\Delta V - 2\beta\rho x_\beta\right)$$

which can be rearranged to give

$$x_\beta = \frac{\frac{E}{A^2}\Delta V}{2\rho\beta\frac{E}{A^2}-1} = \frac{E\Delta V}{2\rho\beta E - A^2}.$$

Substituting E and A from above, and using $e^{\beta\Delta V} = \frac{1-x}{x}$, we have

$$x_\beta = \frac{x(1-x)\Delta V}{2\rho\beta x(1-x)-1}. \tag{57}$$

For $\rho\beta < 2$, the denominator of the right-hand side in (57) is always negative, so $x_\beta < 0$ for $\Delta V > 0$ and $x_\beta > 0$ for $\Delta V < 0$. For $\rho\beta > 2$, the denominator in (57) is negative for $x < x_1^T$ and $x > x_2^T$, with these critical values given in eq. (51), whereas the denominator is positive for $x_1^T < x < x_2^T$. Consequently, the following cases may emerge regarding the sign of x_β:

1. If $\rho\beta \leq 2$, we have:

    - $x_\beta \leq 0$ for $\Delta V > 0$,
    - $x_\beta > 0$ for $\Delta V \leq 0$.

2. If $\rho\beta > 2$, we have:
    - if $0 < x^* < x_1^T$ and $1 > x^* > x_2^T$:
        - for $\Delta V > 0 \Rightarrow x_\beta < 0$;
        - for $\Delta V \leq 0 \Rightarrow x_\beta > 0$;
    - if $0 < x_1^T < x^* < x_2^T < 1$:
        - for $\Delta V < 0 \Rightarrow x_\beta > 0$;
        - for $\Delta V \geq 0 \Rightarrow x_\beta < 0$.

If the flex point is an equilibrium, $\Delta V(x^* = x_F) = 0$, $x_\beta = 0$. Consequently,

$$f'(x^* = x^F) = 2\rho\beta \frac{e^{\beta\Delta V(x_F)}}{\left(1 + e^{\beta\Delta V(x_F)}\right)^2} = \frac{\rho\beta}{2}.$$

In this case, a pitchfork bifurcation occurs at $\beta = 2/\rho$. At this value, the equilibrium $x^* = x^F$ switches from unstable to stable and vice versa.

B.3 The Effect of ρ on x^*

We compute the derivative with respect to ρ on both sides of the fixed point condition for the map in eq. (7). Let $x_\rho = \frac{\partial x^*}{\partial \rho}$, then we have:

$$x_\rho = -\frac{\beta e^{\beta\Delta V}}{\left(1 + e^{\beta\Delta V}\right)^2} \frac{\partial \Delta V}{\partial \rho},$$

and defining $A = \left(1 + e^{\beta\Delta V}\right)$ and $E = e^{\beta\Delta V}$, as $\frac{\partial \Delta V}{\partial \beta} = 1 - 2x - 2\rho x_\rho$, we have

$$x_\rho = \frac{\beta E}{A^2}\left(2\rho x_\rho + 2x - 1\right).$$

Rearranging, we find:

$$x_\rho = \frac{\beta E}{A^2} \frac{1 - 2x}{2\rho\frac{\beta E}{A^2} - 1}.$$

Substituting E and A from above, and using $e^{\beta\Delta V} = \frac{1-x}{x}$, we have:

$$x_\rho = \beta \frac{x(1-x)(1-2x)}{2\rho\beta x(1-x) - 1}. \tag{58}$$

For $\rho\beta < 2$, the denominator of the right-hand side in eq. (58) is always negative. The numerator is positive in $\left[0, \frac{1}{2}\right]$ and negative otherwise. For $\rho\beta > 2$, the denominator is negative for $x < x_1^T$ and $x > x_2^T$, with these critical values given in eq. (51), whereas the denominator is positive for $x_1^T < x < x_2^T$. Consequently, the following cases can arise with respect to the sign of x_ρ.

1. If $\rho\beta \leq 2$, we have:
   - $\frac{\partial x^*}{\partial \rho} \leq 0$ for $0 < x^* < \frac{1}{2}$,
   - $\frac{\partial x^*}{\partial \rho} > 0$ for $1 > x^* > \frac{1}{2}$
2. If $\rho\beta > 2$, there are two characteristic values x_1^T and x_2^T such that
   - if $0 < x^* < x_1^T$ and $1 > x^* > x_2^T$:
     - for $x^* < \frac{1}{2} \Rightarrow x_\rho < 0$;
     - for $x^* > \frac{1}{2} \Rightarrow x_\rho > 0$;
   - if $0 < x_1^T < x^* < x_2^T < 1$:
     - for $x^* < \frac{1}{2} \Rightarrow x_\rho > 0$;
     - for $x^* > \frac{1}{2} \Rightarrow x_\rho < 0$.

If $\lambda = 0$, the flex point $x^F = \frac{1}{2}$ is an equilibrium. The parameter ρ does not affect the position of this equilibrium since $x_\rho = 0$, but it does affect its stability:

$$f'(x^* = x^F = \frac{1}{2}) = 2\rho\beta \frac{e^{\beta\Delta V(x^F=\frac{1}{2})}}{\left(1 + e^{\beta\Delta V(x^F=\frac{1}{2})}\right)^2} = \frac{\rho\beta}{2}.$$

In this case, a pitchfork bifurcation occurs at $\rho = \frac{2}{\beta}$. At this value, the equilibrium $x^* = x^F = \frac{1}{2}$ becomes unstable if it is stable, and vice versa.

Appendix C Proofs of Model Extension I

The effect of the intensity of social interactions ρ on the equilibrium value of x is:

$$\begin{aligned}
\frac{\partial x^*}{\partial \rho} &= \frac{\partial}{\partial \rho}\frac{1}{1 + e^{\beta\Delta V}} \equiv x_\rho \\
&= -\frac{\beta e^{\beta\Delta V}}{\left(1 + e^{\beta\Delta V}\right)^2}\left[(a-1)(1-2x) - 2\rho(a-1)x_\rho\right] \\
&= \beta\frac{E(a-1)}{A^2}\left(1 - 2x - 2\rho x_\rho\right), \qquad (59)
\end{aligned}$$

where $\Delta V = \lambda + \rho(a-1)(1-2x)$, $E = e^{\beta\Delta V}$ and $A = 1 + E$. Rearranging, we get

$$x_\rho = \frac{\beta E(a-1)}{A^2}\frac{1-2x}{2\rho\frac{\beta E(a-1)}{A^2} - 1}. \qquad (60)$$

Since $E = e^{\beta\Delta V} = \frac{1-x}{x}$ and $A = 1 + E = \frac{1}{x}$, we can write

$$\begin{aligned}
x_\rho &= \frac{\beta\frac{1-x}{x}(a-1)}{\frac{1}{x^2}}\frac{1-2x}{2\rho\frac{\beta\frac{1-x}{x}(a-1)}{\frac{1}{x^2}} - 1} \\
&= \frac{\beta(a-1)x(1-x)(1-2x)}{2\rho\beta(a-1)x(1-x) - 1} = (a-1)(1-2x)x_\lambda, \qquad (61)
\end{aligned}$$

where $x_\lambda \equiv \frac{\partial x^*}{\partial \lambda}$. For the effect of the intensity of choice β on the equilibrium, we find:

$$\begin{aligned}\frac{\partial x^*}{\partial \beta} &= \frac{\partial}{\partial \beta}\frac{1}{1+e^{\beta \Delta V}} \equiv x_\beta \\ &= -\frac{\beta e^{\beta \Delta V}}{\left(1+e^{\beta \Delta V}\right)^2}\left[\lambda + \rho(a-1)(1-2x) - 2\rho\beta(a-1)x_\beta\right] \\ &= -\frac{E}{A^2}\left[\lambda + \rho(a-1)(1-2x) - 2\rho\beta(a-1)x_\beta\right]. \end{aligned} \tag{62}$$

Rearranging, we can solve for x_β

$$\begin{aligned} x_\beta &= \frac{[\lambda + \rho(a-1)(1-2x)]x(1-x)}{2\rho\beta(a-1)x(1-x)-1} \\ &= \frac{\Delta V(x)x(1-x)}{2\rho\beta(a-1)x(1-x)-1} = \frac{\Delta V(x)}{\beta}x_\lambda. \end{aligned} \tag{63}$$

Appendix D Proofs of Model Extension II

D.1 Corollary 5.1 and 5.2

Given the variable warm-glow utility term defined in eq. (37), the difference in payoffs is given by

$$\begin{aligned} \Delta V &\equiv V_b - V_g = U(q_b^*) - U(q_g^*) - p^*(q_b^* - q_g^*) + \gamma(q_g^* - \langle q^* \rangle) + \rho(1-2x) \\ &= \zeta + \frac{\gamma(2\delta_g + \gamma)}{2\eta} + \gamma(1-x)\frac{\delta_b - \delta_g - \gamma}{\eta} + \rho(1-2x) \\ &= \zeta' - \gamma(1-x)\frac{\delta_g + \gamma - \delta_b}{\eta} + \rho(1-2x), \end{aligned} \tag{64}$$

where $\zeta = \frac{\delta_g^2 - \delta_b^2}{2\eta}$ and $\zeta' = \zeta + \frac{\gamma(2\delta_g + \gamma)}{2\eta}$.

The first derivative of the function $\Delta V(x)$ with respect to x is given by

$$\Delta V'(x) = -2\rho + \gamma\frac{\delta_g + \gamma - \delta_b}{\eta}.$$

$\Delta V'(x) > 0$ if $\rho < \overline{\rho}$ and $\Delta V'(x) < 0$ if $\rho > \overline{\rho}$, with $\overline{\rho} = \gamma\frac{\delta_g + \gamma - \delta_b}{2\eta}$. Since $\Delta V'(x)$ can also be written as $\Delta V'(x) = -2\rho + 2\overline{\rho}$, by applying the sufficient conditions of eq. (10) in Section 2, this gives directly the conditions in Corollary 5.2.

D.2 The Effects of Parameters on Equilibria and Tipping Points

Using eq. (41) and noticing that $\frac{\partial x}{\partial \zeta} = \frac{\partial x}{\partial \zeta'}\frac{\partial \zeta'}{\partial \zeta}$, we can work either with ζ' or ζ.

$$\frac{\partial x}{\partial \zeta} = -\frac{\beta e^{\beta \Delta V(x)}}{\left(1+e^{\beta \Delta V(x)}\right)^2}\frac{\partial \Delta V(x)}{\partial \zeta} = -\beta x(1-x)\frac{\partial \Delta V(x)}{\partial \zeta}$$

where we have used $x = \frac{1}{1+e^{\beta\Delta V(x)}}$ and $e^{\beta\Delta V(x)} = \frac{1-x}{x}$. Moreover,

$$\frac{\partial \Delta V(x)}{\partial \zeta} = 1 - 2\rho\frac{\partial x}{\partial \zeta} + 2\overline{\rho}\frac{\partial x}{\partial \zeta}$$

such that

$$\frac{\partial x}{\partial \zeta} = -\beta x(1-x) - 2\beta x(1-x)(\overline{\rho} - \rho)\frac{\partial x}{\partial \zeta}$$

in order to derive at

$$\frac{\partial x}{\partial \zeta} = -\frac{\beta x(1-x)}{1 + 2\beta(\overline{\rho} - \rho)x(1-x)}.$$

Since $x(1-x) \leq \frac{1}{4}$, the denominator is always positive for $\rho < \overline{\rho} + \frac{2}{\beta}$, and, hence, $\frac{\partial x}{\partial \zeta} < 0$.

The tipping points of this model are the roots of the denominator of the derivative above:

$$x_1^T = \frac{1 - \sqrt{1 - \frac{2}{\beta(\rho - \overline{\rho})}}}{2}, \quad x_2^T = \frac{1 + \sqrt{1 - \frac{2}{\beta(\rho - \overline{\rho})}}}{2} = 1 - x_1^T.$$

These roots are real numbers if and only if $\rho > \overline{\rho} + \frac{2}{\beta}$, i.e., in a scenario with multiple equilibria. The upper tipping point $1 >^T_2 > \frac{1}{2}$ can be achieved in this scenario through a tangent bifurcation where the green equilibrium disappears. This occurs when ζ increases and achieves the critical value ζ_2^T where $f'(x_2^T) = 1$. Since $x_2^T > \frac{1}{2}$, the numerator of the derivative $\frac{\partial x}{\partial \zeta}$ is negative, such that if the tipping point is reached, the derivative goes to $-\infty$. At the lower tipping point $0 < x_1^T < \frac{1}{2}$, the disappearance of the brown equilibrium occurs at a critical value ζ_1^T where the derivative $\frac{\partial x}{\partial \zeta}$ goes to $+\infty$. Summarising, we can write

- $\lim_{\zeta \to \zeta_2^T} \frac{\partial x^*}{\partial \zeta} = -\infty$, with $\lim_{\zeta \to \zeta_2^T} x^* = x_2^T > \frac{1}{2}$,
- $\lim_{\xi \to \xi_1^T} \frac{\partial x^*}{\partial \xi} = +\infty$, with $\xi = -\zeta$, $\xi_1^T = -\zeta_1^T$ and $\lim_{\xi \to \xi_1^T} x^* = x_1^T < \frac{1}{2}$;

We now turn to the effect of δ_g and δ_b on fixed points $x = f(x)$. (In what follows, we omit the asterisk for fixed points $x^* = f(x^*)$). For δ_g, we have:

$$\frac{\partial x}{\partial \delta_g} = -\beta x(1-x)\frac{\partial \Delta V}{\partial \delta_g}$$

$$\frac{\partial \Delta V}{\partial \delta_g} = \frac{\delta_g}{\eta} - 2\rho\frac{\partial x}{\partial \delta_g} - \frac{\gamma}{\eta}(1-x) + \gamma\frac{\delta_g + \gamma - \delta_b}{\eta}\frac{\partial x}{\partial \delta_g},$$

and, therefore, we have:

$$\frac{\partial x}{\partial \delta_g} = -\beta\frac{\delta_g}{\eta}x(1-x) - \beta x(1-x)\left[\gamma\frac{\delta_g + \gamma - \delta_b}{\eta} - 2\rho\right]\frac{\partial x}{\partial \delta_g} + \beta\frac{\gamma}{\eta}x(1-x)^2.$$

Rearranging, we obtain

$$\begin{aligned}\frac{\partial x}{\partial \delta_g} &= \frac{\frac{\beta}{\eta}x(1-x)[\gamma(1-x)-\delta_g]}{1+\beta\left[\gamma\frac{\delta_g+\gamma-\delta_b}{\eta}-2\rho\right]x(1-x)}\\ \frac{\partial x}{\partial \delta_g} &= \frac{\frac{\beta}{\eta}x(1-x)[\gamma(1-x)-\delta_g]}{1+2\beta(\overline{\rho}-\rho)x(1-x)}\\ &= \frac{\delta_g-\gamma(1-x)}{\eta}\frac{\partial x}{\partial \zeta}.\end{aligned}$$

A sufficient condition for $\delta_g-\gamma(1-x)>0$ is $\delta_g>\gamma$. For δ_b, we have:

$$\frac{\partial x}{\partial \delta_b} = -\beta x(1-x)\frac{\partial \Delta V}{\partial \delta_b},$$

$$\frac{\partial \Delta V}{\partial \delta_b} = -\frac{\delta_b}{\eta}+\frac{\gamma}{\eta}-2\rho\frac{\partial x}{\partial \delta_b}-\frac{\gamma}{\eta}(1-x)+\gamma\frac{\delta_g+\gamma-\delta_b}{\eta}\frac{\partial x}{\partial \delta_b},$$

and, therefore, we get

$$\frac{\partial x}{\partial \delta_b} = \beta\frac{\delta_b}{\eta}x(1-x)-\beta\frac{\gamma}{\eta}x(1-x)+2\rho\beta x(1-x)\frac{\partial x}{\partial \delta_b}-\beta\frac{\gamma}{\eta}x(1-x)^2-\beta\gamma\frac{\delta_g+\gamma-\delta_b}{\eta}\frac{\partial x}{\partial \delta_b}.$$

Rearranging, we obtain

$$\begin{aligned}\frac{\partial x}{\partial \delta_b} &= \frac{\frac{\beta}{\eta}x(1-x)(\delta_b-2\gamma+\gamma x))}{1+\beta\left[\gamma\frac{\delta_g+\gamma-\delta_b}{\eta}-2\rho\right]x(1-x)}\\ \frac{\partial x}{\partial \delta_b} &= \frac{\frac{\beta}{\eta}x(1-x)(\delta_b-2\gamma+\gamma x)}{1+2\beta(\overline{\rho}-\rho)x(1-x)}\\ &= \frac{2\gamma-\gamma x-\delta_b}{\eta}\frac{\partial x}{\partial \zeta}.\end{aligned}$$

A sufficient condition for $2\gamma-\gamma x-\delta_b<0$ is $\delta_b>2\gamma$ and for $2\gamma-\gamma x-\delta_b>0$ a sufficient condition is $\delta_b<\gamma$. For γ, we have:

$$\frac{\partial x}{\partial \gamma} = \beta x(1-x)\left[2(\rho-\overline{\rho})\frac{\partial x}{\partial \gamma}+\frac{\delta_g-\delta_b+2\gamma}{\eta}(1-x)-\frac{\delta_g+\gamma}{\eta}\right],$$

which can be rewritten as

$$\begin{aligned}\frac{\partial x}{\partial \gamma} &= \frac{\beta x(1-x)}{\eta}\frac{(\delta_g-\delta_b+2\gamma)(1-x)-\delta_g-\gamma}{1+2\beta x(1-x)(\overline{\rho}-\rho)} =\\ &= -\frac{(\delta_g-\delta_b+2\gamma)(1-x)-\delta_g-\gamma}{\eta}\frac{\partial x}{\partial \zeta}.\end{aligned}$$

The numerator of the first factor is equal to $\gamma(1-2x)-(\delta_g-\delta_b)x-\delta_b$. This expression is clearly negative for $x>\frac{1}{2}$. Hence, for $x>\frac{1}{2}$, the sign of $\frac{\partial x}{\partial \gamma}$ is equal to the sign of $\frac{\partial x}{\partial \zeta}$.

Given that we are considering a green equilibrium ($x>\frac{1}{2}$), only the second tipping point exists and maybe reached with an increase of γ. A critical (bifurcation) value of γ exists such that

$$\lim_{\gamma\to\gamma^T}\frac{\partial x^*}{\partial \gamma} = -\infty, \quad \text{with} \quad \lim_{\gamma\to\gamma^T}x^* = x^T > \frac{1}{2}.$$

Appendix E Social Welfare

Social welfare is given by

$$SW(x) = \sum_{i=1}^{n} CS_i(q_i^*(x)) + PS(Q^*(x)) - \sum_{i=1}^{n} D_i(Q^*(x)),$$

which can be rewritten as

$$SW(x) = \sum_{i=1}^{n} U_i(q_i^*(x)) - PC(Q^*(x)) - \sum_{i=1}^{n} D_i(Q^*(x)),$$

because expenses of consumers and revenues of producers cancel out. In all versions of our model, the utility from material consumption $U_i(q_i^*(x))$ is given by eq. (12) with $i = g, b$, and aggregate damages $\sum_{i=1}^{n} D_i(Q^*(x)) = nD(\sum_{i=1}^{n} q_i^*) = n\delta Q^*(x)$ follow from eq. (13). Moreover, the assumption is that aggregate supply is a linear function given by (18). Hence, we assume production costs to be given by $PC(Q^*(x)) = \frac{1}{2s}(Q^*)^2$, even though any other affine transformation would lead to the same qualitative conclusion. In the basic model and in extension I, individual equilibrium quantities are given by (16) and the aggregate quantity by (17). Using these quantities as inputs to calculate social welfare, straightforward but cumbersome calculations show that $SW(x)$ is a U-shaped function because

$$\frac{\partial^2 SW(x)}{\partial x^2} = \frac{(\delta_g - \delta_b)^2 n^2}{(\eta s + n)\eta} > 0,$$

with a minimum at

$$x_{min} = \frac{(\delta_g + \delta_b)\eta s + (\delta_g - \delta_b)n - 2\delta n s \eta}{2n(\delta_g - \delta_b)}.$$

By assumption $\delta_g > \delta_b$, and $SW(x)$ is increasing in $[0, 1]$ (i.e., $x_{min} < 0$) for

$$\frac{(\delta_g + \delta_b)\eta s + (\delta_g - \delta_b)n}{2ns\eta} = \frac{(\delta_g + \delta_b)}{2n} + \frac{(\delta_g - \delta_b)}{2s\eta} \equiv \underline{\delta} < \delta,$$

which we assume to hold.

For extension II, hardly anything changes. We use eq. (38) for individual consumption, and (39) for aggregate consumption, and obtain

$$\frac{\partial^2 SW(x)}{\partial x^2} = \frac{(\delta_g + \gamma - \delta_b)^2 n^2}{(\eta s + n)\eta} > 0,$$

and $x_{min} < 0$ if

$$\frac{(\delta_g + \gamma + \delta_b)\eta s + (\delta_g + \gamma - \delta_b)n}{2ns\eta} = \frac{(\delta_g + \gamma + \delta_b)}{2n} + \frac{(\delta_g + \gamma - \delta_b)}{2s\eta} \equiv \underline{\underline{\delta}} < \delta.$$

The message is that "objective", or "realised" marginal damage δ needs to be sufficiently large for total welfare to increase in x in the entire domain $[0, 1]$.

Appendix F Effect of Parameters on Consumption Patterns, Total Damages and Global Welfare

The following analysis uses the results of Section 3.2, in particular eqs. (16) to (19), about equilibrium quantities and prices, and of Lemma 3.1 and Proposition 3.3 about the effect of parameters on the equilibrium fraction of x of individuals adopting a green lifestyle.

F.1 Basic Model: the Effect of γ, ρ and β (Proposition 6.1)

The effect of γ on the equilibrium price $p^*(x)$ of eq. (19) is given by:

$$\frac{\partial p^*}{\partial \gamma} = -\frac{\delta_g - \delta_b}{1 + \eta s/n} \frac{\partial x^*}{\partial \gamma} = -\frac{\delta_g - \delta_b}{1 + \eta s/n} \frac{\partial x^*}{\partial \lambda} \frac{\partial \lambda}{\partial \gamma} = \frac{\delta_g - \delta_b}{1 + \eta s/n} \frac{\partial x^*}{\partial \lambda}$$

where we have used $\frac{\partial \lambda}{\partial \gamma} = -1$, as $\lambda = \zeta - \gamma$. From Lemma 3.1 we know that $\frac{\partial x^*}{\partial \lambda} < 0$ if x^* is a stable equilibrium and no tipping point. Thus, $\frac{\partial x^*}{\partial \gamma} > 0$. Consequently, $\frac{\partial p^*}{\partial \gamma} < 0$. Because $Q^*(x) = sp^*(x)$, we have $\frac{\partial Q^*}{\partial \gamma} < 0$ whereas $\frac{\partial q_i^*}{\partial \gamma} > 0$ for $i = g, b$. At a tipping point, if γ increases, λ decreases. Therefore, if the equilibrium is a brown equilibrium at the social tipping point, it will disappear such that the population will move to the green equilibrium so that Q^* decreases in a non-continuous way.

The effect of β on $p^*(x)$ is given by:

$$\frac{\partial p^*}{\partial \beta} = -\frac{\delta_g - \delta_b}{1 + \eta s/n} \frac{\partial x^*}{\partial \beta} = -\frac{\delta_g - \delta_b}{1 + \eta s/n} \frac{\Delta V}{\beta} \frac{\partial x^*}{\partial \lambda}.$$

Since $\frac{\partial x^*}{\partial \lambda} < 0$ in a stable equilibrium, we have $\frac{\partial p^*}{\partial \beta} > 0$ in a brown equilibrium ($\Delta V > 0$) and $\frac{\partial p^*}{\partial \beta} < 0$ in a green equilibrium ($\Delta V < 0$). Because of $Q^*(x) = sp^*(x)$, equilibrium price and aggregate quantity change in the same direction, whereas individual demand changes in the opposite direction (according to eq. (16)). As an increase in β reinforces equilibria but does not tip them, nothing changes if x^* is a tipping point.

The effect of ρ on $p^*(x)$ is given by:

$$\frac{\partial p^*}{\partial \rho} = -\frac{\delta_g - \delta_b}{1 + \eta s/n} \frac{\partial x^*}{\partial \rho} = -\frac{\delta_g - \delta_b}{1 + \eta s/n} (1 - 2^* x) \frac{\partial x^*}{\partial \lambda}.$$

Since $\frac{\partial x^*}{\partial \lambda} < 0$ in a stable equilibrium, we have $\frac{\partial p^*}{\partial \rho} > 0$ in a brown equilibrium ($x^* < \frac{1}{2}$) and $\frac{\partial p}{\partial \rho} < 0$ in a green equilibrium ($x^* > \frac{1}{2}$). The total equilibrium quantity changes in the same direction and individual quantities in the opposite direction. As for changes in β, an increase of ρ will not tip equilibria.

We notice that total environmental damages are a function of Q^* (which is a function of x as spelled out above) and social welfare is an increasing function of x^* as derived in Appendix E.

F.2 Basic Model: the Effect of δ_b and δ_g (Propositions 6.2 and 6.3)

For the effect of the internalisation parameter of browns δ_b, we have:

$$\begin{aligned}\frac{\partial p^*}{\partial \delta_b} &= -\frac{1}{1+\eta\frac{s}{n}} + \frac{x^*}{1+\eta\frac{s}{n}} - \frac{\delta_g - \delta_b}{1+\eta\frac{s}{n}}\frac{\partial x^*}{\partial \delta_b} \\ &= -\frac{1}{1+\eta\frac{s}{n}}\left(1 - x^* + (\delta_g - \delta_b)\frac{\partial x^*}{\partial \delta_b}\right).\end{aligned}$$

We have $\frac{\partial x^*}{\partial \delta_b} = \frac{\partial x^*}{\partial \lambda}\frac{\partial \lambda}{\partial \delta_b} = -\frac{\delta_b}{\eta}\frac{\partial x^*}{\partial \lambda} > 0$, as $\frac{\partial x^*}{\partial \lambda} < 0$ if $x^* \notin [x_1^T, x_2^T]$, i.e., when x^* is a stable equilibrium not at a social tipping point, as explained in Appendix B.1. Consequently, the term in brackets is positive. Hence, we have:

$$\frac{\partial p^*}{\partial \delta_b} < 0 \quad \text{and} \quad \frac{\partial Q^*}{\partial \delta_b} < 0.$$

For the effect of individual consumption levels of greens, we therefore have:

$$\frac{\partial q_g^*}{\partial \delta_b} = -\frac{1}{\eta}\frac{\partial p^*}{\partial \delta_b} > 0.$$

For browns, we find:

$$\frac{\partial q_b^*}{\partial \delta_b} = -\frac{1}{\eta}\left(1 + \frac{\partial p^*}{\partial \delta_b}\right).$$

Thus, we can distinguish between two cases.

i) If $-1 < \frac{\partial p^*}{\partial \delta_b} < 0$, then $\frac{\partial q_b^*}{\partial \delta_b} < 0$.

ii) If $\frac{\partial p^*}{\partial \delta_b} < -1$, then $\frac{\partial q_b^*}{\partial \delta_b} > 0$.

At the tipping point x_1^T, the brown equilibrium suddenly disappears if δ_b increases and Q^* drops abruptly.

For social welfare, we have:

$$\frac{\partial SW(x^*, \delta_b)}{\partial \delta_b} = \frac{\partial SW}{\partial x^*}\frac{\partial x^*}{\partial \delta_b} + \frac{\partial SW}{\partial \delta_b}.$$

δ_b affect social welfare via a change of x^*, which is the indirect effect, and via a change of consumption levels, which is the direct effect. We know that $\frac{\partial SW}{\partial x^*} > 0$ if the "objective" damage parameter δ is sufficiently large, i.e. $\delta > \underline{\delta} \equiv \frac{(\delta_g+\delta_b)\eta s + (\delta_g - \delta_b)n}{2ns\eta}$, as derived in Appendix E. We also know from Proposition 3.2 that $\frac{\partial x^*}{\delta_b} > 0$ if x^* is not a tipping point and if it is a tipping point the brown equilibrium will disappear so that this derivative becomes infinite. Finally, we compute

$$\frac{\partial SW}{\partial \delta_b} = \frac{(1-x)n\left[(\delta_g - \delta_b)\, nx + \delta n s \eta - \delta_b s \eta\right]}{\eta(\eta s + n)} \tag{65}$$

(for any given x) for which a sufficient condition that this derivative is positive is $\delta n s \eta - \delta_b s \eta > 0$, or $\delta > \delta_b / n$, which does not seem a strong assumption. Thus, all three components in $\frac{\partial SW(x^*, \delta_b)}{\partial \delta_b}$ are positive.

Regarding the effect of the internalisation parameter of greens, δ_g, we have:

$$\frac{\partial p^*}{\partial \delta_g} = -\frac{x^*}{1 + \eta \frac{s}{n}} - \frac{\delta_g - \delta_b}{1 + \eta \frac{s}{n}} \frac{\partial x^*}{\partial \delta_g}.$$

We have $\frac{\partial x^*}{\partial \delta_g} = \frac{\partial x^*}{\partial \lambda} \frac{\partial \lambda}{\partial \delta_g} = \frac{\delta_g}{\eta} \frac{\partial x^*}{\partial \lambda} < 0$, as $\frac{\partial \lambda}{\partial \delta_g} = \frac{\partial \zeta}{\partial \delta_g} = \frac{\delta_g}{\eta}$ and $\frac{\partial x^*}{\partial \lambda} < 0$ if $x^* \notin [x_1^T, x_2^T]$.

The first term of $\frac{\partial p^*}{\partial \delta_g}$ on the right-hand side is negative. This is the direct effect of decreasing consumption of greens due to a higher degree of internalisation δ_g. The second term on the right-hand side is positive. This is the indirect or composition effect due to a smaller fraction of greens. A higher degree of internalisation implies that the difference in utility derived from brown versus green consumption increases, making the green lifestyle less attractive, i.e, x^* drops. Accordingly, we can distinguish two cases.

i) If $\left| \frac{\partial x^*}{\partial \delta_g} \right| < \frac{x^*}{\delta_g - \delta_b}$, then $\frac{\partial p^*}{\partial \delta_g} < 0$ and also $\frac{\partial Q^*}{\partial \delta_g} < 0$.

ii) If $\left| \frac{\partial x^*}{\partial \delta_g} \right| > \frac{x^*}{\delta_g - \delta_b}$, then $\frac{\partial p^*}{\partial \delta_g} > 0$ and also $\frac{\partial Q^*}{\partial \delta_g} > 0$.

In case (i), the direct effects is stronger than the indirect effect, in case (ii), this is reversed. For individual consumption levels, we find

$$\frac{\partial q_b^*}{\partial \delta_g} = -\frac{1}{\eta} \frac{\partial p^*}{\partial \delta_g}.$$

This will be positive in case (i) and negative in case (ii), and always has the opposite sign than equilibrium total consumption. Moreover, we have

$$\frac{\partial q_g^*}{\partial \delta_g} = -\frac{1}{\eta} - \frac{1}{\eta} \frac{\partial p^*}{\partial \delta_g} = -\frac{1}{\eta} \left(1 + \frac{\partial p^*}{\partial \delta_g} \right).$$

Thus, we can distinguish two cases.

a) If $\frac{\partial p^*}{\partial \delta_g} > -1$, then $\frac{\partial q_g^*}{\partial \delta_g} < 0$. This maybe true in case (i), and is always true in case (ii) above.

b) If $\frac{\partial p^*}{\partial \delta_g} < -1$ then $\frac{\partial q_g^*}{\partial \delta_g} > 0$. This maybe true in case (i), but never holds in case (ii) above.

Considering tipping points, an increase in δ_g increases λ. Thus, a green equilibrium x_g^* could tip to a brown equilibrium x_b^*, which, not only in case (ii), leads to higher total consumption, but this could (and most likely will) also happen in case (i), which de facto becomes case (ii), as $\left| \frac{\partial x^*}{\partial \delta_g} \right|$ becomes suddenly very large and x^* very small.

For social welfare, we have:

$$\frac{\partial SW(x^*,\delta_g)}{\partial \delta_g} = \frac{\partial SW}{\partial x^*}\frac{\partial x^*}{\partial \delta_g} + \frac{\partial SW}{\partial \delta_g}.$$

We know that $\frac{\partial SW}{\partial x^*} > 0$ if δ is sufficiently large. From Proposition 3.2, $\frac{\partial x^*}{\delta_g} > 0$ if x^* is not a tipping point, and if x^* is a tipping point, the green equilibrium will disappear (through a bifurcation, as the derivative is minus infinity). Moreover, we have

$$\frac{\partial SW}{\partial \delta_g} = \frac{nx\left[(\delta_g - \delta_b)(nx-1) + \delta n s\eta - \delta_g s\eta\right]}{\eta(\eta s + n)}. \tag{66}$$

This is positive, if $\delta n s\eta - \delta_g s\eta > 0$, which gives $\delta > \delta_g/n$ and $nx - 1 > 0$, i.e. $x > 1/n$. Both conditions are mild conditions. Nevertheless, because the indirect effect is negative and the direct effect is positive, nothing can be generally concluded. At a tipping point, social welfare will most likely decrease as $\frac{\partial x^*}{\partial \delta_g}$ becomes infinitely negative.

References

Andreoni, J. (1990): "Impure Altruism and Donations to Public Goods: a Theory of Warm-Glow Giving," *Economic Journal*, 464–477.

——— (1995): "Warm-glow versus Cold-prickle: the Effects of Positive and Negative Framing on Cooperation in Experiments," *Quarterly Journal of Economics*, 110, 1–21.

Baiocchi, G., J. Minx, and K. Hubacek (2010): "The impact of social factors and consumer behavior on carbon dioxide emissions in the United Kingdom," *Journal of Industrial Ecology*, 14, 50–72.

Bell, P. A., J. D. Fisher, A. Baum, and T. C. Greene, eds. (1990): *Environmental psychology*, Fort Worth, TX, USA: Holt, Rinehart and Winston.

Bicchieri, C. (2006): *The Grammar of Society*, Cambridge, England: Cambridge University Press.

Brekke, K., S. Kverndokk, and K. Nyborg (2003): "An economic model of moral motivation," *Journal of Public Economics*, 87, 1967–1983.

Brock, W. and S. Durlauf (2001): "Discrete choice with social interactions," *Review of Economic Studies*, 68, 235–260.

Brock, W. and C. Hommes (1997): "A rational route to randomness," *Econometrica*, 65, 1059–1095.

Cialdini, R. B. and N. J. Goldstein (2004): "Social influence: Compliance and conformity," *Annual Review of Psychology*, 55, 591–621.

Hommes, C. (2013): *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*, Cambridge, England: Cambridge University Press.

Kuznetsov, Y. (1998): *Elements of applied bifurcation theory*, New York: Springer.

Manski, C. (2000): "Economic analysis of social interactions," *Journal of Economic Perspectives*, 14, 115–136.

McFadden, D. L. (1981): "Structural discrete probability models derived from theories of choice," in *Structural analysis of discrete data and econometric applications*, ed. by C. Manski and D. L. McFadden, Cambridge, US: MIT Press.

Medio, A. and G. Gallo (1995): *Chaotic dynamics*, Cambridge, England: Cambridge University Press.

Ryan, R. M. and E. L. Deci (2000): "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *American psychologist*, 55, 68.

Van der Linden, S. (2018): "Warm glow is associated with low-but not high-cost sustainable behaviour," *Nature Sustainability*, 1, 28–30.

Zeppini, P. (2015): "A discrete choice model of transitions to sustainable technologies," *Journal of Economic Behavior and Organization*, 112, 187–203.

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2026
GREDEG Working Papers Released in 2026

2026-01	Estelle Malavolti & Frédéric Marty *L'équilibre économique des aéroports secondaires européens à l'épreuve des dynamiques de concurrence fiscale*
2026-02	Charlie Joyez *A New Index of Export-Import Proximity: Conceptual Foundations and Global Patterns*
2026-03	Patrice Bougette & Frédéric Marty *The SCP Paradigm Revisited: What Structuralism Really Contributed to U.S. Antitrust*
2026-04	Imen Bouhlel, Nathalie Lazaric & Paolo Zeppini *Competitive Diffusion and Sustainability Transitions: The Case of Plastics Recycling Technologies*
2026-05	Maxime Menuet *Fractional Replicator Dynamics*
2026-06	Ilona Dielen, Patrice Bougette & Christophe Charlier *A (Green) Switch in Time Saves Nine: Assessing the Environmental Damage of the European Truck Cartel*
2026-07	Thibault Schrepel & Godefroy de Boiscuillé *Questioning the Digital Markets Act's Legality*
2026-08	Michael Finus & Paolo Zeppini *Green Lifestyles and Social Tipping Points*