Citation for published version:
Wainwright, A 2020 'A Model of Pro-Environmental Behaviour with Heterogeneous Agents'.

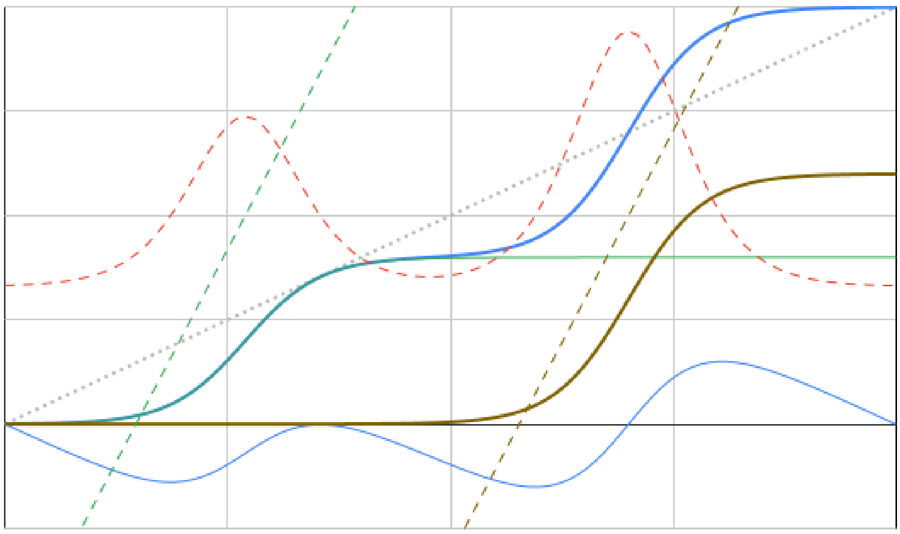
Publication date:
2020

Link to publication

A Model of Pro-Environmental Behaviour with Heterogeneous Agents

Andrew Wainwright

24th September 2020



Contents

1 Introduction

Anthropogenic climate change poses an existential threat to our society, at a scale not seen for at least 4 million years (Feldmann and Levermann, 2015; Foster, Royer and Lunt, 2017). Emissions of greenhouse gasses (mostly carbon dioxide, methane and nitrous oxides) have caused levels of CO_2 in the atmosphere to reach levels no human being has ever witnessed before. We are currently well into the planet's sixth mass extinction (Barnosky et al., 2011).

Faced with overwhelming evidence of the issues, people around the world are increasingly demanding that their governments respond appropriately (L.C. Hamilton et al., 2015; Leiserowitz et al., 2018; Prieur, 2020; Tong et al., 2019; UNFCCC, 2015). We are however, still waiting for the action (den Elzen et al., 2019; Ge et al., 2019; Masson-Delmotte et al., 2018).

It is possible we may have a long wait. It has been argued that people care insufficiently and that climate change is someone else's problem (see for example C. Hamilton, 2015; Hasselmann et al., 2003; Marshall, 2015). We do know that there are many reasons why people can choose to act pro-socially, even at great cost to themselves (Bicchieri, 2006; Thøgersen, 2008; Viscusi, Huber and Bell, 2011).

It is therefore very important that we improve our ability to model the predictors of pro-environmental behaviour, so that we are better placed to design policy interventions which enhance our ability to address the current crisis.

Current models of pro-environmental behaviour are limited in their ability to predict the effect of social norms. This paper takes a small step in addressing that by taking an existing model and extending it.

We introduce an additional level of heterogeneity to a model of pro-environmental behaviour based on discrete choice. We find that this gives rise to an additional equilibrium which has a good fit with "real-world" experiences.

The goal of this paper is to explore the current model in some depth, so that it can be used as a yard-stick against which to compare the new model. In doing so we bring some new insights to the existing model as well as the new one.

Section 2 reviews the base model, largely drawing on the work of Brock and Durlauf; Zeppini (2001; 2015). We establish a number of properties of the model, and are able to find necessary and sufficient conditions for the existence of multiple equilibria.

Having established a base for comparison, section 3 then adds an additional level of heterogeneity to the model, by admitting two distinct sets of preferences. We examine the effect this has on the behaviour of individuals and the dynamics of the model.

2 Base Model

This section looks at a model of innovation using discrete choice from Zeppini (2015). We take the time to examine its characteristics in some detail, for subsequent comparison with section 3.

We start with a statement of the scope of the model—a number of individuals with a binary choice to make, and utility arising from either option. We introduce social norms as an element of the utility function, modelled here as a two-sided linear function of the "popularity" of the choice—the proportion of people making the same choice.

We then refer to the discrete choice framework set up by Brock and Durlauf (2001) and in particular by Zeppini (2015). Under the conditions in the model, the resulting distribution is given by a logistic function, and we briefly recap on the features of this curve, with attention to the implications for this model (and that of section 3).

The model is then completed by the addition of the requirement to repeat the choice indefinitely. This allows us to ask questions of the dynamics of the model, and to look for the existence of equilibria (stable and unstable) in the models. We close the section with a classification of solution types.

Section 3 then follows a similar path but applied to the heterogeneous version of the model.

2.1 Utility Function

A population $\boldsymbol{P}$ of utility-maximising individuals is required to choose between two options which we label "sustainable" and "unsustainable." Each option gives the individual a base utility (W_j, for $j \in \{s, u\}$). Without loss of generality we assert $W_u \geq W_s$, and we define $\Delta W = W_u - W_s$.

Let ω_i be a choice variable, defined for each individual $i \in \{1, 2, \ldots, N\}$:

$$\omega_i = \begin{cases} s, & \text{if individual } i \text{ chooses the } \textit{sustainable} \text{ option} \\ u, & \text{if individual } i \text{ chooses the } \textit{unsustainable} \text{ option.} \end{cases} \tag{1}$$

We define N as the size of $\boldsymbol{P}$ and $x \in [0, 1]$ as the proportion which choose the *sustainable* option:

$$N = |\boldsymbol{P}| \tag{2}$$

$$x = \frac{\Sigma_{i=1}^{N} \mathbb{1}_{\omega_i = s}}{N} \tag{3}$$

In addition to W_j, each individual also receives utility $S_i \in \{S_s, S_u\}$ from a descriptive social norm, which is to say that they derive additional utility if the option they choose is popular. See figure 1.

$$S_i = S_{\omega_i} = \begin{cases} S_s = \rho x, & \text{if } \omega_i = s \\ S_u = \rho(1 - x), & \text{if } \omega_i = u. \end{cases} \tag{4}$$

$\rho > 0$ is a fixed strength parameter governing the strength of social interactions. If $\rho = 0$ then the social utility is effectively "turned off." This results in a model which is independent of x: the probability of an individual choosing the *sustainable* option is a constant, dependent on ΔW. We therefore exclude this case from our model.

S_i increases with the proportion x of people making the *sustainable* choice, but also with the proportion of people making an *unsustainable* choice $(1 - x)$, as shown in figure 1. If $x < \frac{1}{2}$, then they will be obtaining higher social utility if they have chosen the *unsustainable* option. If they have chosen the *sustainable* option, then the social utility they obtain is less than they would have received had they chosen the other option. So $S_s > S_u$ if $x > \frac{1}{2}$ and $S_u > S_s$ if $x < \frac{1}{2}$.

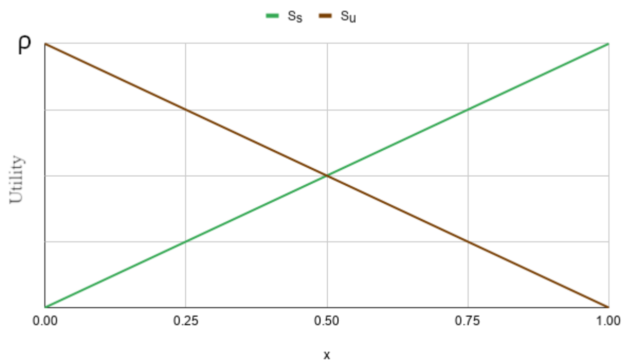


Figure 1: Utility given by the social norm function. Individuals receive utility given by *either* the green (S_S) line, *or* the brown (S_u) line, depending on their ω_i.

2.2 Loose Preferences

We say that an individual "loosely prefers" a given option if that option has higher base utility W. Just because they may loosely prefer one particular option though does not mean they will necessarily choose that option: if the social factor is high enough they will opt "against" their preference.

It is instructive to ask under what conditions individuals will do this, and we will look at this in section 2.5 below. For now, note that if individuals "loosely prefer" the *unsustainable* option, they will only go against this preference if the social utility they would obtain outweighs ΔW, and this can only be the case if $x > \frac{1}{2}$. This makes sense intuitively: if the *unsustainable* option is the one an individual loosely prefers (that is, it has the higher base utility), then it will take positive social "pressure" for the individual to choose the *sustainable* option, and S_s is only greater than S_u when $x > \frac{1}{2}$.

2.3 Discrete Choice

We can model individual behaviour using a discrete choice framework (see Brock and Durlauf, 2001; Manski and McFadden, 1981; Zeppini, 2015). Discrete choice introduces a stochastic "noise shock" (ε) to the utility function, which has the effect of introducing a certain heterogeneity into the model. Individuals each have their own ε_i. Without this, all individuals would choose the same option at any given value of x. ε can be thought of as representing variation in decision making between individuals (due for instance to bounded rationality or else to exogenous psychological factors). Each individual's ε is independent of any other individual's ε and they are all identically distributed with variance σ^2.

The individual's utility function can now be written in full:

$$U_i = W_{\omega_i} + S_{\omega_i} + \varepsilon_i \tag{5}$$

Following Brock and Durlauf, we obtain a distribution of responses, given by

$$\begin{aligned} Pr\{\text{individual } i \text{ chooses the } \textit{sustainable} \text{ option}\} &\equiv y \\ &= \frac{e^{\beta U_s}}{e^{\beta U_u} + e^{\beta U_s}} = \frac{1}{1 + e^{\beta(\Delta W + \rho(1-2x))}} \end{aligned} \tag{6}$$

where $\beta \in [0, \infty)$ is a parameter which can be interpreted as the "intensity of choice," inversely proportional to σ^2, the variance of ε.

Figure 2 shows the effect of β on the distribution. Increasing the intensity of choice (β) increases the gradient of the curve in its steepest part. In the limit (as $\beta \to \infty$) the graph becomes a step function, with the "noise" reduced to zero, as every individual makes the same "rational" decision given x. Conversely, when $\beta = 0$, then the probability an individual chooses either option is $\frac{1}{2}$, for all values of ρ or ΔW. It is as if there is so much noise inherent in ε that individuals' choices simply cannot be predicted.

Also shown on figure 2 are the first derivatives of the two curves. These are characteristic of their primitives and provide a more immediate way to see both variables of the logistic curve: the location of the flex point and the maximum gradient. (It is easier for the eye to extract height information than a gradient, and the location of a point is more evident when there is a curve reaching a maximum than to try and judge when a curve is most steep.)

An item of notation: in this paper we find it convenient to write:

$$f(x, \beta, \Delta W, \rho) = f(x) = y = \frac{1}{1 + e^{\beta(\Delta W + \rho(1-2x))}} \tag{7}$$

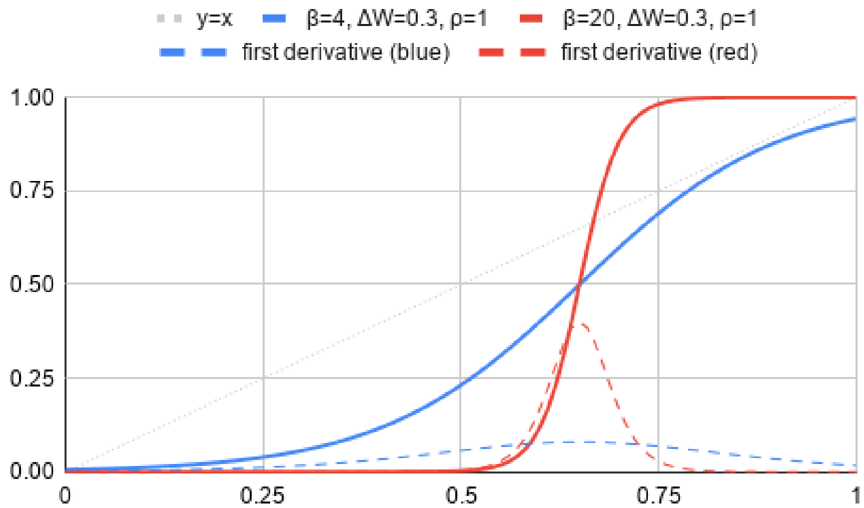


Figure 2: Two probability distributions of an individual's choice as a function of x, the proportion of individuals choosing the *sustainable* option. The blue curve has low β; the red cure has identical ΔW and ρ, but a higher β. Notice how the two curves intersect only at the flex point. Also shown (in dashed lines) are the first derivatives of each curve (on a compressed axis).

2.4 Logistic Function

Equation 7 is a logistic curve, and logistic curves have a number of useful features. Restricting ourselves to the subfamily of increasing logistic curves we can write:

$$y = \frac{\kappa}{1 + e^{-a(x-b)}}\ , \quad \text{with } a > 0, \ \ b, \kappa \in \mathbb{R} \tag{8}$$

and make the following claims:

Proposition 2.1

The curve has asymptotes at $y = 0$ and $y = \kappa$.

Proof. $x \in (-\infty, \infty) \Rightarrow e^{-a(x-b)} \in (0, \infty)$, so the range of y is $(0, \kappa)$. $e^{-a(x-b)}$ is monotonic decreasing so y is monotonic increasing. Thus y is monotonic and bounded at $y = 0$ and $y = \kappa$. $\square$

Proposition 2.2

The curve has a single flex point at $x = b$.

Proof. Using the reciprocal rule,

$$\frac{dy}{dx} = \frac{a\kappa e^{-a(x-b)}}{(1+e^{-a(x-b)})^2} > 0$$

From the quotient rule,

$$\frac{d^2y}{dx^2} = \frac{a^2\kappa(e^{-a(x-b)} - 1)\, e^{-a(x-b)}}{(1+e^{-a(x-b)})^3}$$

which changes sign exactly once, when $e^{-a(x-b)} = 1 \iff x = b$. □

Proposition 2.3

The value of y at the flex point is $\frac{\kappa}{2}$.

Proof.

$$y = \frac{\kappa}{1+e^{-a(b-b)}} = \frac{\kappa}{1+e^0} = \frac{\kappa}{2}$$

□

Corollary 2.4

The maximum gradient is reached at the flex point, and is equal to $\frac{a\kappa}{4}$.

Corollary 2.5

$\frac{dy}{dx}$ is positive for the interval $(-\infty, b)$ and negative for the interval (b, ∞).

Proposition 2.6

The curve has rotational symmetry around its flex point, so that $y(z) = \kappa - y(-z),$ *where* $z = x - b$.

Proof.

$$\kappa - y(-z) = \frac{\kappa(1 + e^{-a(-z)}) - \kappa}{1 + e^{-a(-z)}} = \frac{\kappa e^{az}}{1 + e^{az}} = \frac{\kappa}{e^{-az} + 1} = y(z)$$

$\square$

Comparing equations 6 and 8, we obtain the following parametrisation:

$$\begin{aligned} \kappa = &\text{ upper bound for probability } y &= 1 \\ a = &\text{ maximum gradient times 4} &= 2\beta\rho \\ b = &\text{ location of flex point} &= \tfrac{1}{2} + \tfrac{\Delta W}{2\rho} \end{aligned} \tag{9}$$

and we may refer to $f(x, \beta, \Delta W, \rho)$ as $f(x, a, b)$. We will also label the point $x_I \equiv b$. At this point, $y = \frac{1}{2}$, so this is precisely the point at which the individual i has a fifty percent chance of choosing either option - the *indifference* point. It is easy to find graphically: it is simply the point where the curve crosses the $y = \frac{1}{2}$ line. It is also easy to compute analytically ($\frac{1}{2} + \frac{\Delta W}{2\rho}$).

Note that the indifference point may not exist $\in [0, 1]$. If ΔW is very large (or ρ very small) then the indifference point $= \frac{1}{2} + \frac{\Delta W}{2\rho} > 1$. The intuition behind this is an individual saying "I don't care if everyone else in the world disagrees; I'm not spending that much extra money." If the social pressure (ρ) were higher, or if the extra cost were lower (ΔW), then their statement might change. So for the indifference point to be *feasible*, we require:

$$\begin{aligned} & x_I \leq 1 \\ \iff & \tfrac{1}{2} + \tfrac{\Delta W}{2\rho} \leq 1 \\ \iff & \frac{\Delta W}{\rho} \leq 1 \end{aligned} \tag{10}$$

We note also that the maximum gradient is independent of ΔW, and that the location of the indifference point is independent of β. ρ affects both the gradient and the indifference point: an increase in ρ results in an increase in the maximum gradient, but also moves the indifference point to the left.

The maximum steepness of the curve is proportional not just to the intensity of choice (β), but also to ρ, the strength of social interactions. We shall return to this point in section 2.5 later.

2.5 Dynamics

To finally complete the model, the choice is repeated, indefinitely. $x = x_t$ is then a function of discrete time, with a given starting value x_0. We can use equation 7 as the revision protocol, and write

$$x_t = \frac{1}{1 + e^{\beta(\Delta W + \rho(1 - 2x_{t-1}))}} \equiv f(x_{t-1}) \tag{11}$$

We are naturally interested in how x changes over time, and in the thresholds between various states. The first question is whether equilibria are possible, and if so, how many, and of what type. Because x is a proportion, it is constrained to the range $[0, 1]$. The constraint on x means that we are only interested in equilibria that may occur in that range. This allows us to make certain statements concerning limits to the dynamics.

We will show:

- That there is at least one and at most three equilibria;
- If there are three equilibria x_1, x_2, x_3 with $x_1 < x_2 < x_3$, then $x_1 < x_I$ and $x_3 > x_I$ where x_I is the flex point (indifference point);
- If there are three equilibria then x_1 and x_3 are stable equilibria and x_2 is an unstable equilibrium;
- If there are two equilibria, then one is stable, and the other is semi-stable (stable in one direction only);
- If there is one equilibrium then it is stable;
- If $x_I > \frac{1}{2}$ then $x_2 > x_I$ and if $x_I < \frac{1}{2}$ then $x_2 < x_I$.

We are also able to make statements concerning the conditions required for the existence of three equilibria.

Graphical Example

First however, it may be useful to look at an example (see figure 3). This curve has $\beta = 3$, $\Delta W = 0.8$ and $\rho = 1$, and has three equilibria in $[0, 1]$. The location of the equilibria (x_1, x_2 and x_3) is shown by the vertical green bars.

Notice how iteration of $x_{t+1} = f(x)$ from the initial brown and green starting points gives values which are further away from the x_2 equilibrium. In contrast, iteration in the neighbourhood of x_1 or x_3 results in convergence towards those equilibria.

We find it useful to define $g(x, \beta, \Delta W, \rho) = f(x, \beta, \Delta W, \rho) - x$, and to use the shorthand $g(x) = f(x) - x$. Then points of equilibrium in the model correspond to roots of the equation $g(x) = 0$.

We can also clearly see that what determines whether an equilibrium is stable or an unstable is the gradient of $g(x)$ at the point of equilibrium. If this gradient is negative, as it is at x_1 and x_3, then the equilibrium is stable; if the gradient is positive (as at x_2) then the equilibrium is unstable.

The equilibria define four zones. In zone A ($x < x_1$), $x < f(x) < x_1$, so x will increase and converge on x_1.

In zone B, ($x_1 < x < x_2$), $x_1 < f(x) < x$, so x will decrease and converge on x_1 (shown by the brown arrows).

In zone C ($x_2 < x < x_3$), $x < f(x) < x_3$, so x will increase and converge on x_3 (shown by the purple arrows).

In zone D, ($x > x_3$), $x_3 < f(x) < x$, so x will decrease and converge on x_3.

We can now explore whether the curve always has these features, and how many equilibria we can expect to find.

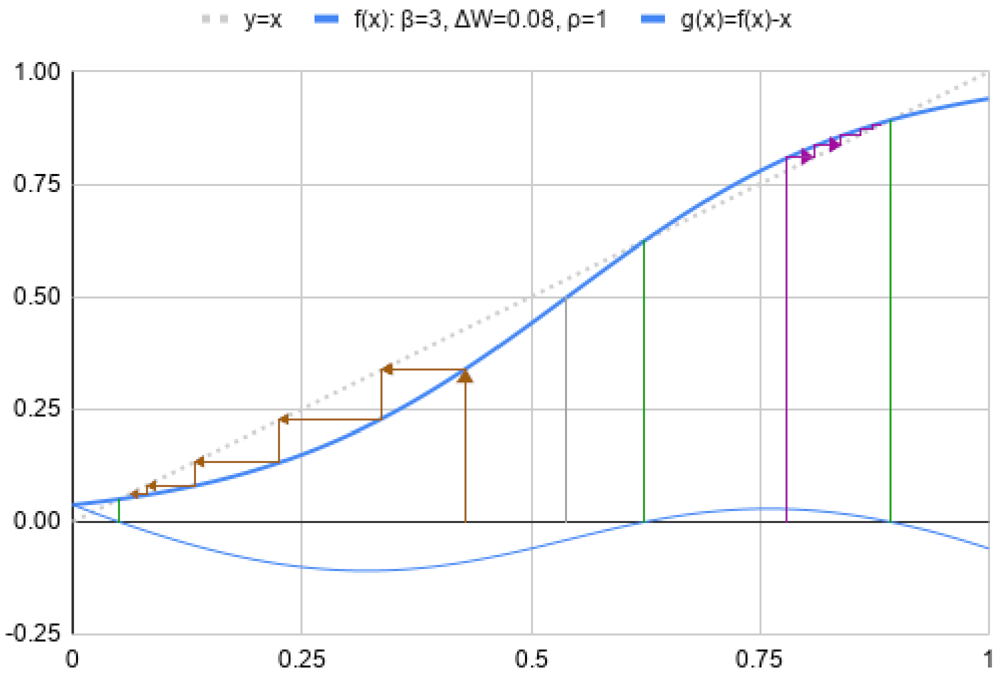


Figure 3: Iteration towards the attractors near $x = 0$ and $x = 1$. The green vertical bars mark the three equilibria, and the grey vertical bar shows the location of x_I. The brown and purple paths show the result of iterating $x_{t+1} = f(x_t)$ from below and above (respectively) the flext point. The thinner blue line shows $g(x) = f(x) - x$, and makes it easier to see when $f(x)$ is above or below the $y = x$ line.

Proposition 2.7

There is at least one equilibrium.

Proof. For a value x^* to represent an equilibrium, we need $x^* = f(x^*)$, so we are looking for roots of the equation

$$g(x) \equiv f(x) - x = 0 \tag{12}$$

$g(x)$ is a continuous and differentiable function with $g(0) = \frac{1}{1+e^{\beta(\Delta W+\rho)}} > 0$ for all values of β, ΔW and ρ. Similarly, $g(1) = \frac{1}{1+e^{\beta(\Delta W+\rho)}} - 1 < 0$. Then by the mean value theorem, there must exist at least one $x^* \in (0, 1)$ such that $g(x^*) = 0 \iff f(x^*) = x^*$. $\square$

Proposition 2.8

There can be no more than three equilibria.

Proof. The second derivative,

$$\frac{d^2g}{dx^2} = \frac{(2\beta\rho)^2(e^{\beta(\Delta W+\rho(1-2x))}-1)e^{\beta(\Delta W+\rho(1-2x))}}{(1+e^{\beta(\Delta W+\rho(1-2x))})^3}$$

changes sign only once, when

$$\begin{aligned} & e^{\beta(\Delta W+\rho(1-2x))} = 1 \\ \iff & \Delta W + \rho(1-2x) = 0 \\ \iff & x = \tfrac{1}{2} + \tfrac{\Delta W}{2\rho}, \end{aligned}$$

so the highest number of roots of $g(x)$ is three. $\square$

Proposition 2.9

If there are three equilibria x_1, x_2, x_3 with $x_1 < x_2 < x_3$, then $x_1 < x_I$ and $x_3 > x_I$, where x_I is the indifference point.

Proof. Three equilibria means that $g(x)$ goes from positive to negative, back to positive and then back to negative, in the range $[0,1]$. Suppose there is no equilibrium in the range $[0,x_I)$. Then there must be three in the range $[x_I,1]$. However by corollary 2.5, $f(x)$ is monotonic decreasing in this range, and as $f(x_I) > x$ there can be only one equilibrium in that range—a contradiction. Therefore there is at least one equilibrium in the range $[0,x_I)$. By a similar argument, there must be at least one equilibrium in the range $(x_I,1]$. $\square$

Proposition 2.10

If there are three equilibria then x_1 and x_3 are stable equilibria and x_2 is unstable.

Proof. At both x_1 and x_3, $g(x)$ goes from positive to negative, so the gradient at these two points is (locally) negative. In these regions, $x_{t+1} = f(x_t)$ will be closer to the equilibrium than x_t. At x_2, $g(x)$ goes from negative to positive, so the gradient is locally positive, and subsequent iterations of x_t will be further away from x_2. □

Proposition 2.11

If there are two equilibria, then one is stable and the other is semi-stable.

Proof. $g(0) > 0$ and $g(1) < 0$, so for there to be two equilibria, one of the two must be a "glancing" equilibrium, where $\frac{dg}{dx} = 0$ exactly at the point where $g(x) = 0$. Type 2a has the "glancing" equilibrium first, so that $x_1 = x_2$. Type 2b has the "glancing" equilibrium second ($x_2 = x_3$). See figure 6 which has examples of both types. In Type 2a, the equilibrium at x_3 is exactly as in the three equilibrium case: $g(x)$ goes from positive to negative, so the gradient at these two points is (locally) negative, making it a stable equilibrium. The other equilibrium (at $x_1 = x_2$) is more interesting. $g(x)$ is positive on both sides of the point of equilibrium, which means that successive iterations of x converge *towards* the point for $x < x_1$, but *away* for $x > x_1$. Type 2b is the same in reverse: the equilibrium at x_1 is the stable one, and the one at $x_2 = x_3$ is semi-stable, with the stable side being $x > x_3$. □

Proposition 2.12

If there is only one equilibrium then it is stable.

Proof. $g(0) > 0$ and $g(1) < 0$, so there is an $x_1 \in (0, 1)$ such that $g(x_1) = 0$. For values of x near this point, $g(x) > 0$ for $x < x_1$ and $g(x) > 0$ for $x > x_1$. Thus successive iterations of x_t will converge towards this equilibrium, making it stable. □

Notice that no matter how many equilibria there are, there is always at least one stable one, and iterations starting from the extreme values (0 and 1) will converge away from these extremes.

Proposition 2.13

If $x_I > \frac{1}{2}$ *then* $x_2 > x_I$ *and if* $x_I < \frac{1}{2}$ *then* $x_2 < x_I$*.*

Proof.

$$\begin{aligned} & x_I > \tfrac{1}{2} \\ \iff \quad & \tfrac{1}{2} + \tfrac{\Delta W}{2\rho} > 0 \\ \iff \quad & \Delta W > 0, \end{aligned}$$

as $\rho > 0$. Then $f(x_I) = \frac{1}{2}$ (by proposition 2.3). But $x_I > \frac{1}{2}$, so $g(x_I)$ is still negative and $x_2 > x_I$. Conversely, if $x_I < \frac{1}{2}$ then $\Delta W < 0$, and $g(x_I)$ is already positive, so $x_2 < x_I$. □

We can also go some way in pinning down the conditions for existence of three equilibria:

Proposition 2.14

A necessary condition for the existence of three equilibria is that $\beta\rho > 2$*.*

Proof. By proposition 2.8, x_1 exists, and $g(x) < 0$ for $x > x_1$. But if $\beta\rho \leq 2$ then the maximum gradient of the logistic curve, which is $\frac{a}{4}$ by proposition 2.2, is less than or equal to 1. Therefore for the interval $(x_1, 1]$ $\frac{dg}{dx} \leq 1$, and $g(x) \leq x$. □

Proposition 2.15

A necessary condition for the existence of three equilibria is that $\frac{\Delta W}{\rho} < 1$.

Proof (by contradiction). Assume there are three equilibria and $\Delta W \geq \rho$. Then $x_I = \frac{1}{2} + \frac{\Delta W}{2\rho} \geq 1$. By corollary 2.5, it follows that $f(x)$ is monotonic increasing throughout the range $[0, 1]$, and $f(0) > 0$ and $f(1) < 1$, so there can only be one equilibrium in this case, a contradiction. □

Compare this result with equation 10 on page 13, which required $\frac{\Delta W}{\rho} \leq 1$ in order that the indifference point fall in the range $[0, 1]$. Proposition 2.15 therefore requires the existence of an indifference point as a pre-requisite for the curve having three equilibria.

The previous two propositions addressed necessity, but it is also possible to give *sufficient* conditions for the existence of three equilibria. We start by considering a special case:

Proposition 2.16

The indifference point is an equilibrium if and only if $\Delta W = 0$.

Proof.
$$\begin{aligned} & \Delta W && = 0 \\ \iff \quad & x_I && = \tfrac{1}{2} + \tfrac{\Delta W}{2\rho} = \tfrac{1}{2} \quad (\rho > 0) \\ \iff \quad & f(x) && = \tfrac{1}{2} \quad \text{(by proposition 2.2)} \\ & && = x \end{aligned}$$

(The uniqueness of the point at which $f(x) = \frac{1}{2}$ is because $f(x)$ is monotonic.)

□

Proposition 2.17

If $\beta\rho > 2$ *and* $\Delta W = 0$*, then* $f(x, \beta, \Delta W, \rho)$ *has three equilibria.*

Proof. $y = f(x)$ intersects the line $y = x$ at the indifference point, and at this point (by corollary 2.4) $\frac{df}{dx} = \frac{\beta\rho}{2} > 1$. We also have $g(x) < 0$ for $x < x_I$ and $g(x) > 0$ for $x > x_I$. Which means $g(x)$ goes from positive (at $x = 0$) to negative (just below x_I), back to positive (just above x_I) and then back to negative again (at $x = 1$) so $f(x, \beta, 0, \rho)$ must have three equilibria. □

Proposition 2.14 established the necessity of $\beta\rho > 2$ (that is, $a/4 > 1$); without this, the curve $y = f(x)$ never becomes steep enough for $g(x) = f(x) - x$ to change sign sufficiently often. What is needed in addition is for the line $y = x$ to intersect the logistic curve close to the point where the sign changing is happening—in other words, close to the indifference point.

Proposition 2.18

If $\beta\rho > 2$ then there exists a $\delta \in (0, \frac{1}{2})$ such that if $|\frac{\Delta W}{\rho}| < \delta$ then $f(x, \beta, \Delta W, \rho)$ has three equilibria.

Proof. Consider the equation $y = f(x, a, b)$ and let b vary, treating y as a function of b. $\frac{\partial y}{\partial b} < 0$ for the whole of the $[0, 1]$ range, so increasing b has the effect of shifting both x_1 and x_2 to the right. However, $\frac{\partial y}{\partial b}$ is higher at the flex point than at the extremes, so x_2 moves faster than both x_1 and x_3. The result is that the interval $x_2 - x_1$ increases and the interval $x_3 - x_2$ decreases, until it reaches zero, at which point $g(x, a, b)$ ceases to have three roots. If the value of b at which x_2 meets x_3 is b^+, then we can also define b^- by decreasing b until x_1 meets x_2. Then $\delta = b^+ - \frac{1}{2}$. By proposition 2.6, $b^+ - \frac{1}{2} = \frac{1}{2} - b^-$ so for $|\frac{\Delta W}{\rho}| < \delta$, $b \in (b^-, b^+)$ and $g(x, \beta, \Delta W, \rho)$ has three roots. □

Finally, we are able to provide a value for δ.

Proposition 2.19

The value of δ for $f(x, \beta, \delta W, \rho)$ is equal to the maximum value of $g(x, \beta, 0, \rho)$ evaluated between x_2 and x_3.

Proof. Let x_H be the point in the range (x_1, x_2) at which $g(x)$ reaches its maximum.

$$\left.\frac{dg}{dx}\right|_{x_H} = 0$$

This is well-defined because

- $g(x)$ is continuous and differentiable throughout $[0, 1]$,
- $g(x_2) = g(x_3) = 0$,
- $x_2 \neq x_3$,
- $f(x) > 0$ for all $x \in (x_2, x_3)$
- and $\frac{d^2g}{dx^2} = \frac{d^2y}{dx^2}$ so by proposition 2.2 $g(x)$ has no points of inflection for $x < b = x_I = x_2$.

We can similarly define $x_L \in (x_1, x_2)$ as the point at which:

$$\left.\frac{dg}{dx}\right|_{x_L} = 0$$

Note that by proposition 2.6, $\frac{1}{2} - x_L = x_H - \frac{1}{2}$ and $g(x_L) = -g(x_H)$.

Let $d \equiv g(x_H) = f(x_H, a, b = 0) - x_H$.

We now transform coordinates so that $z = x + \delta$, for any $\delta \in (0, d)$.

$$\begin{aligned} f(x, a, b = 0) = \frac{1}{1 + e^{-ax}} &= \frac{1}{1 + e^{-a(z-\delta)}} \\ &= f(z, a, b = \delta) \\ &= f(x + \delta, a, b = \delta) \end{aligned}$$

but at x_H, $f(x_H, a, b = 0) = x_H + d$ so

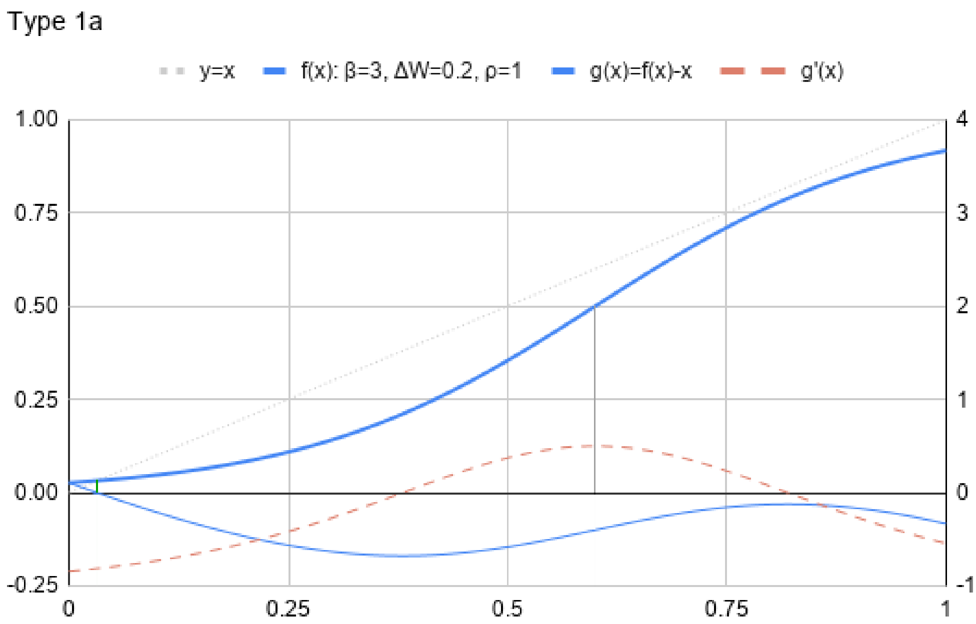
$$f(x_H, a, b = 0) = f(x_H + \delta, a, b = \delta) = x_H + d > x_H + \delta.$$

so $f(x + \delta, a, b = \delta)$ also has three equilibria. □

It is not possible to calculate solutions of this model analytically. However, various computational methods can be used without difficulty. The examples detailed in the figures used in this paper sometimes have solutions calculated via the Newton-Raphson method, for instance (Newton, 1687).

2.6 Classification

We can now classify instances of this model by the number (and type) of equilibria they give rise to; see figures 4, 5 and 6.



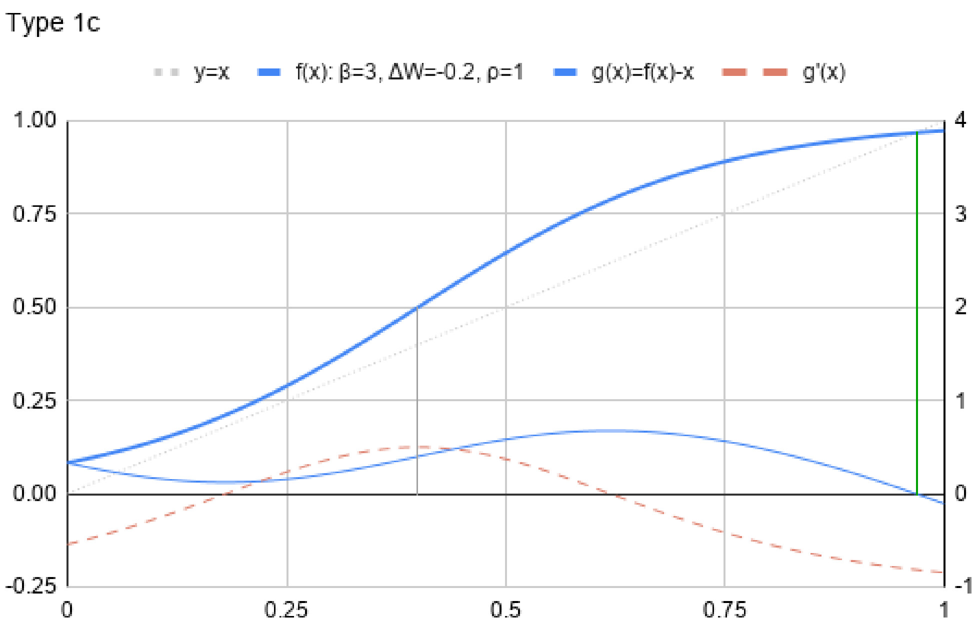
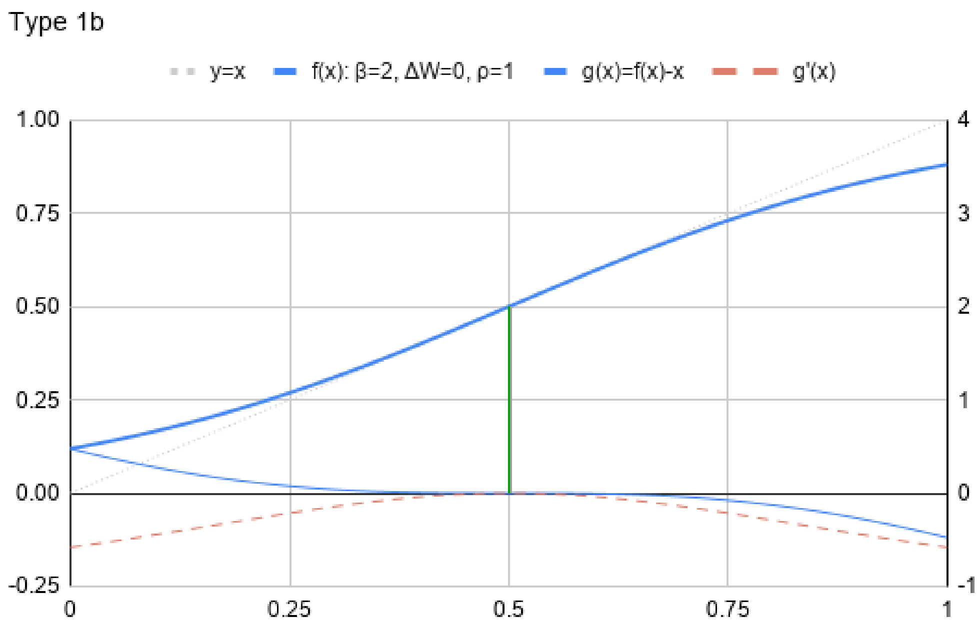


Figure 4: Logistic curve classification: type 1a (upper), and 1c (lower). For type 1b please see figure 5. Both these types (and type 1b) have one stable equilibrium. In type 1a, $x_1 < x_I$; type 1c has $x_1 > x_I$; The green vertical lines indicate the equilibria and the grey vertical lines show the location of the flex point x_I. The right vertical axes are for the dashed red line $g'(x) = \frac{dg}{dx}$ and the left vertical axis is for all other series.



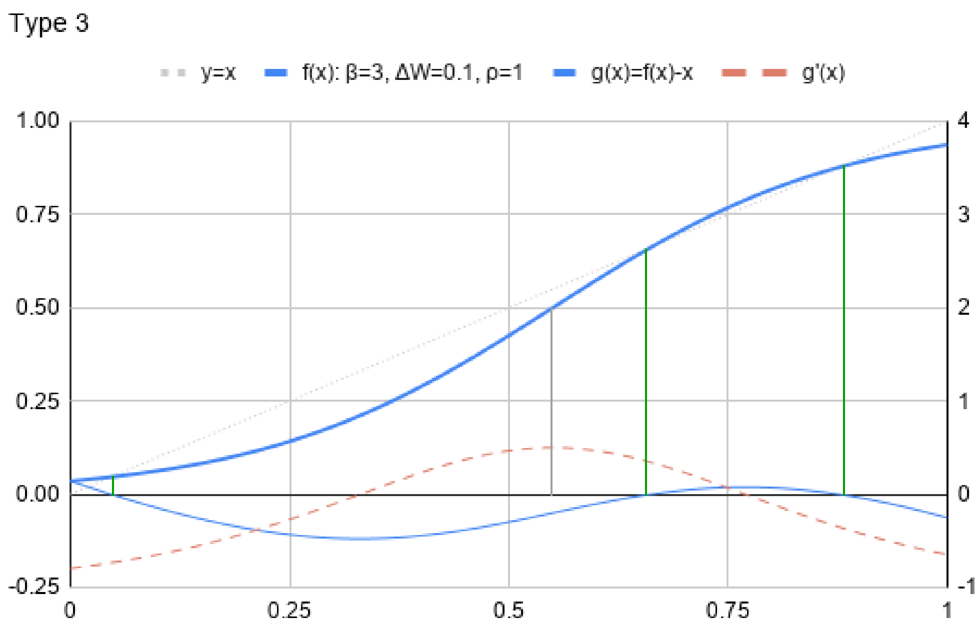


Figure 5: Logistic curve classification: type 1b (upper), and 3 (lower). For types 1a and 1c please see figure 4. Type 1b has one stable equilibrium, type 3 has two stable equilibria and one unstable equilibrium. In type 1b, $x_1 = x_I$; type 3 has $x_1(stable) < x_I < x_3(stable)$; The unstable equilibrium (x_2) can fall either side of x_I (in this example, $x_2 > x_I$). The green vertical lines indicate the equilibria and the grey vertical lines show the location of the flex point x_I. The right vertical axes are for the dashed red line $g'(x) = \frac{dg}{dx}$ and the left vertical axis is for all other series.

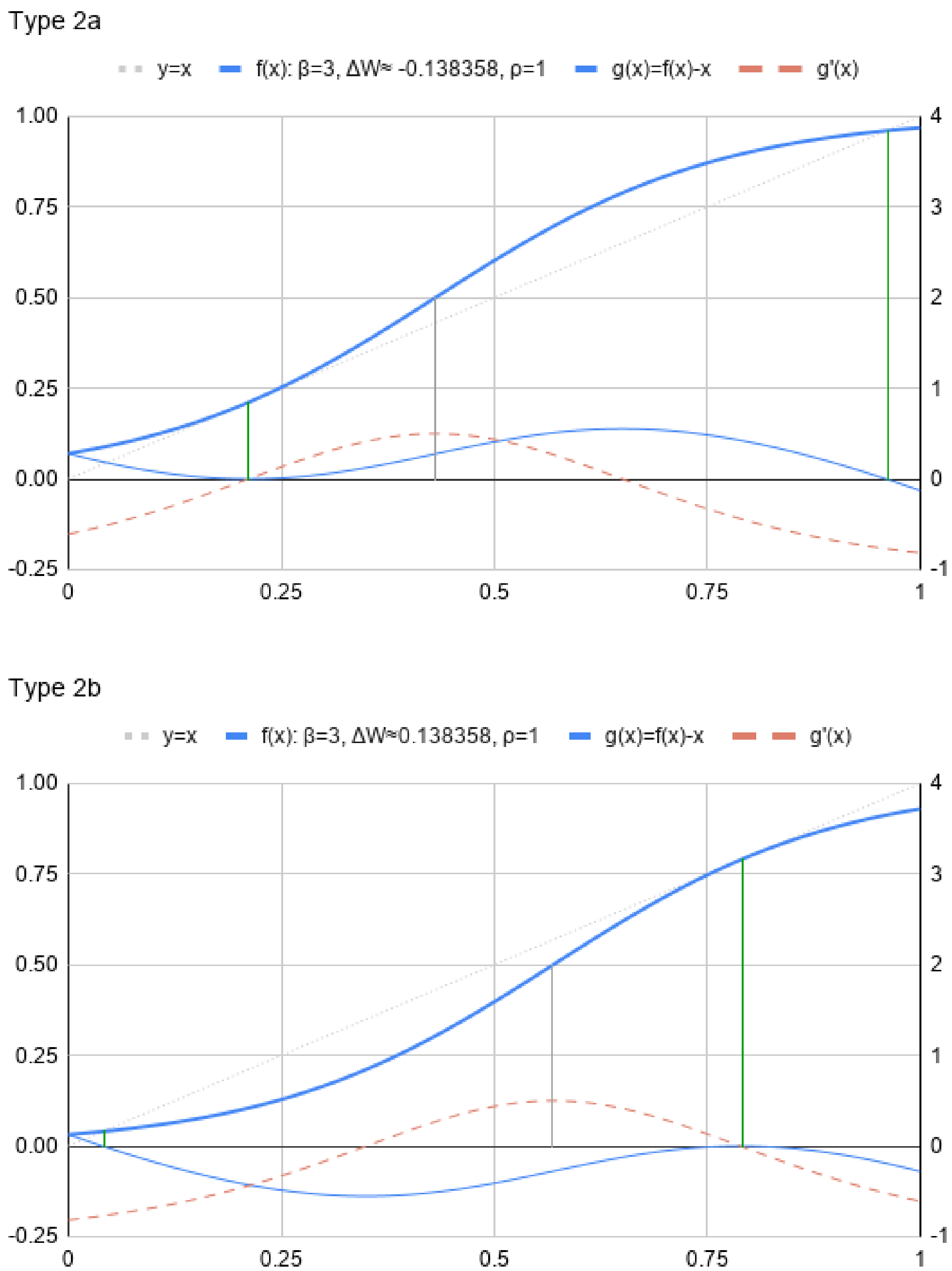


Figure 6: Logistic curve classification: type 2a (upper), and 2b (lower). Both these types have two equilibria. In type 2a, $x_1 < x_I$; type 1c has $x_1 > x_I$; The green vertical lines indicate the equilibria and the grey vertical lines show the location of the flex point x_I. The right vertical axes are for the dashed red line $g'(x) = \frac{dg}{dx}$ and the left vertical axis is for all other series.

2.7 Summary

There are, in this base model, only two stable equilibria, and therefore only two potential outcomes. Either the population converges the equilibrium near zero, which corresponds to almost all of the individuals deciding on the *unsustainable* option, or else on the equilibrium near 1, which corresponds to the *sustainable* option.

The dynamics are as follows: if the proportion of the population choosing the *sustainable* option lies below a certain threshold ($\frac{1}{2} + \frac{\Delta W}{2\rho}$) then that proportion will continue to drop, heading towards zero. If the proportion is greater than $\frac{1}{2} + \frac{\Delta W}{2\rho}$, then the population will increasingly choose the *sustainable* option.

If the *sustainable* option is unpopular, then the majority of the individuals will choose the *unsustainable* option. However if the *sustainable* option is popular, then their choice will depend on how much more utility they gain from the social norm. If this utility is greater than the relative disutility from choosing the *sustainable* option, then they will choose the *sustainable* option.

Figure 7 shows the regions of the $(x,\, \rho)$ space in which the *sustainable* and the *unsustainable* options are the attractors.

Let us now see how this model is affected by the addition of heterogeneous agents.

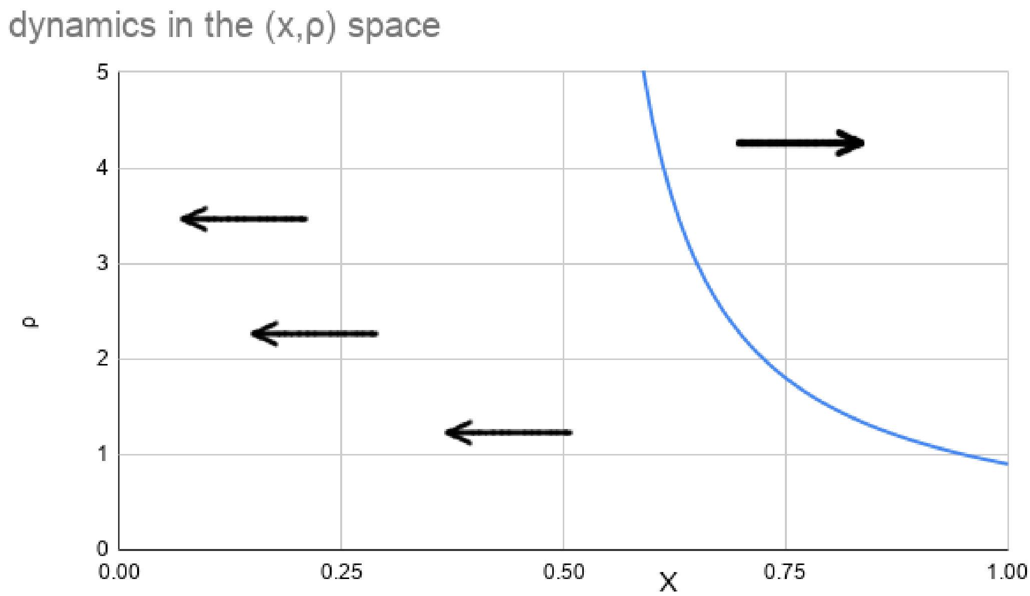


Figure 7: Regions of the $(x,\, \rho)$ space and the attractors in operation there.

3 Heterogeneity

We now introduce the heterogeneous model, in which individuals are of one of two types, which we will call *environmentalist* (*env*) and $non - environmentalist$ (*non*) depending on whether they get higher base utility from *sustainable* choices.

Instead of the two different possible values of base utility in the simple model, we now have four: *environmentalist* (*env*) individuals get W_s^{env} utils for a *sustainable* choice and W_u^{env} for an *unsustainable* one. Similarly, $non - environmentalist$ (*non*) individuals get W_s^{non} and W_u^{non} utils.

Let $\boldsymbol{E} \subseteq \boldsymbol{P}$ be the subset of the population which is *environmentalist*, and $\boldsymbol{N} = \boldsymbol{P} \setminus \boldsymbol{E}$ those who are $non - environmentalist$. We define $\mu \in [0, 1]$ as the proportion of *environmentalist* individuals in the population:

$$\mu = \frac{|\boldsymbol{E}|}{|\boldsymbol{P}|}, \quad \text{so } \frac{|\boldsymbol{N}|}{|\boldsymbol{P}|} = 1 - \mu. \tag{13}$$

3.1 Utility Function

Recalling our definition of "loose preference" from section 2.2, an *environmentalist* "loosely prefers" the *sustainable* option but a $non - environmentalist$ "loosely prefers" the *unsustainable* one. This gives us $W_s^{env} \geq W_u^{env}$ and $W_u^{non} \geq W_s^{non}$. We extend the definition of ΔW:

$$\Delta W^t \equiv W_u^t - W_s^t, \qquad \text{for } t \in \{env, non\}.$$

Note that although $\Delta W^{non} \geq 0$, we now have $\Delta W^{env} \leq 0$. This is a departure from the base model; ΔW can now be negative (at least for *environmentalist* individuals, that is, $i \in \boldsymbol{E}$). Comparing choices in the heterogeneous model, individuals are indifferent between the *sustainable* and *unsustainable* options when

$$\begin{aligned} W_s^t + \rho x_I^t + \varepsilon_i &= W_u^t + \rho(1 - x_I^t) + \varepsilon_i \\ \iff x_I^t &\equiv \tfrac{1}{2} + \tfrac{\Delta W^t}{2\rho} \end{aligned} \tag{14}$$

So for *environmentalist* individuals, their indifference point is now less than $\frac{1}{2}$, which means they require more than 50% of the population to be choosing the *unsustainable* option before they themselves will choose it.

Now that ΔW can be negative, we have an additional feasibility requirement. For x_I^{env} to exist $\in [0, 1]$,

$$\begin{aligned} \tfrac{1}{2} + \tfrac{\Delta W^{env}}{2\rho} &\geq 0 \\ \iff -\Delta W^{env} &\leq \rho \end{aligned}$$

that is,

$$W_s^{env} - W_u^{env} \leq \rho$$

Combining this with the feasibility requirement on ΔW^{non} carried forward from the base model, we can summarise both restrictions:

$$\frac{|\Delta W^t|}{\rho} \leq 1, \quad \text{for } t \in \{env, non\}$$

Utility now depends on the individual's type:

$$U_i = W^t_{\omega_i} + S_{\omega_i} + \varepsilon_i = \begin{cases} W^{env}_{\omega_i} + S_{\omega_i} + \varepsilon_i, & \text{if } i \in \boldsymbol{E} \\ W^{non}_{\omega_i} + S_{\omega_i} + \varepsilon_i, & \text{if } i \in \boldsymbol{N} \end{cases} \tag{15}$$

3.2 Discrete Choice

The proportion of individuals choosing the *sustainable* option is now given as:

$$\begin{aligned} x_t &= Pr\{\omega_i = sus\} \\ &= Pr\{\omega_i = sus | i \in \boldsymbol{E}\}.Pr\{i \in \boldsymbol{E}\} + Pr\{\omega_i = sus | i \in \boldsymbol{N}\}.Pr\{i \in \boldsymbol{N}\} \\ &= \frac{\mu}{1 + e^{\beta(\Delta W^{env} + \rho(1-2x_{t-1}))}} + \frac{1-\mu}{1 + e^{\beta(\Delta W^{non} + \rho(1-2x_{t-1}))}} \\ &= \mu f(x_{t-1}, \beta, \Delta W^{env}, \rho) + (1-\mu) f(x_{t-1}, \beta, \Delta W^{non}, \rho) \\ &\equiv c(\mu, x_{t-1}, \beta, \Delta W^{non}, \rho) \end{aligned} \tag{16}$$

This is a weighted sum of two logistic functions, forming in the general case a quasi-sigmoidal curve with an additional "step," which may be more or less prominent, depending on the precise parameters[1]. See figure 8, which shows the (red) curve compared to the $y = x$ line. Parameters have been chosen in this graph so that (in addition to the attractors near $x = 0$ and $x = 1$) there are three other equilibria: two unstable ones and a stable one between them. Depending on parameters, some of these equilibria may not exist.

[1] This has been described as the shape obtained when carpet is laid over two stairs without stapling.

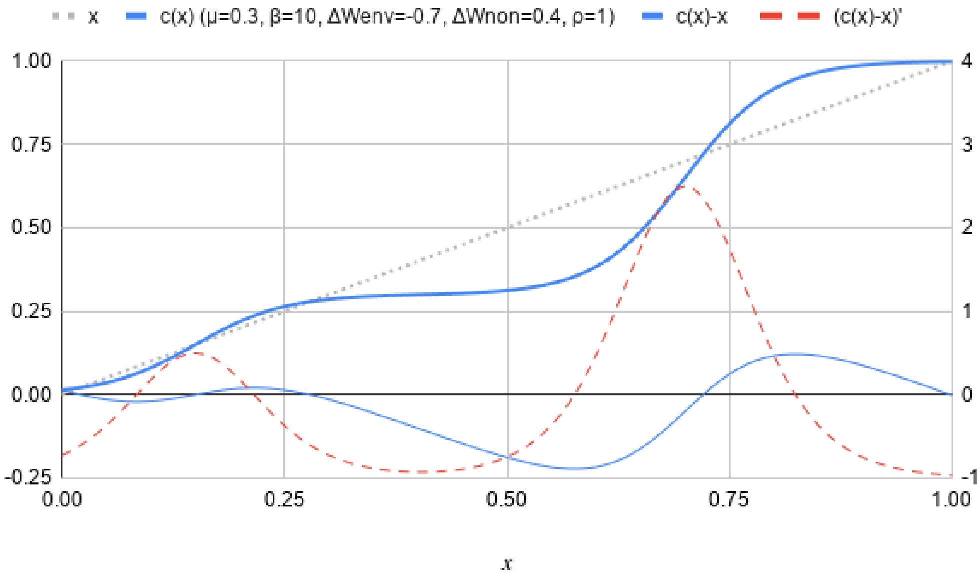


Figure 8: Bi-logistic function $c(x)$ in the thick blue line. Also shown on the same axis $c(x) - x$ (thinner blue line) and $y = x$ in the dotted grey line. The first derivative $\frac{d(c(x)-x)}{dx}$ is shown in the dashed red line against the right-hand axis. The points of equilibrium of $c(x)$ are the roots of the equation $c(x)-x = 0$ and the indifference points coincide with the two local maxima on the dashed red line.

3.3 Linear Representation

Fisher and Pry (1971) showed how logistic curves can be transformed into straight lines, from which the values of a and b can easily be obtained (Fisher and Pry, 1971). The method is particularly useful for the analysis of empirical data based on these models, but also serves a purpose here. They first define

$$F = \frac{y}{\kappa}$$

and then plot $\frac{F}{1-F}$ against x with a logarithmic ordinate (G).

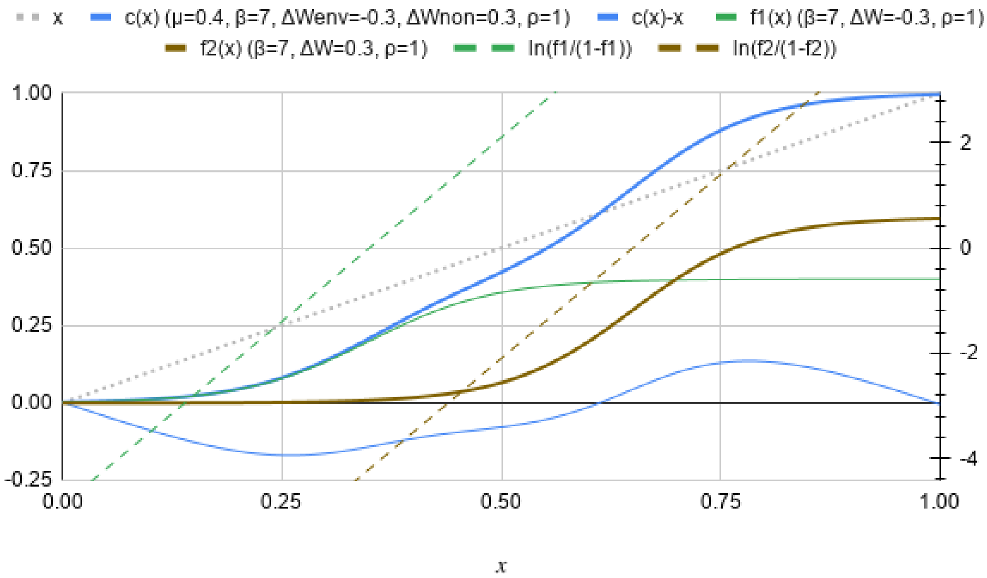
$$G \equiv \frac{F}{1-F} = \frac{y}{1-y} = \frac{\frac{1}{1+e^{-a(x-b)}}}{\frac{e^{-a(x-b)}}{1+e^{-a(x-b)}}} = \frac{1}{e^{-a(x-b)}} = e^{ax-ab} \tag{17}$$

Alternatively, $ln(G)$ can be plotted against x, which is what we have done in figure 9. This transformation yields a straight line which is characteristic of the original logistic curve. As can be seen from equation 17, the gradient of the line is a and the intercept is $-ab$.

Another useful feature of the Fisher-Pry transform is that when $y = \frac{1}{2}$, $ln\left(\frac{y}{1-y}\right) = 0$. It makes sense therefore, when plotting graphs of curves against two different axes, to align them so that the $y = \frac{1}{2}$ line on one axis corresponds to the $ln\left(\frac{y}{1-y}\right) = 0$ line on the other. There is then a degree of choice about how to scale the (Fisher-Pry) axis. In figure 9 we have chosen to scale so that $y = 0$ and $y = 1$ on one axis correspond to $ln\left(\frac{0.05}{0.95}\right)$ and $ln\left(\frac{0.95}{0.05}\right)$ on the Fisher-Pry axis. These are the points where the logistic curves have reached 5% and 95% (respectively) of their full value.

Figure 9 shows the Fisher-Pry linearisation of the logistic curves which comprise the two bi-logistic curves $c(x, \mu = 0.4, \boldsymbol{\beta = 7}, \Delta W^{env} = -0.3, \Delta W^{non} = 0.3, \rho = 1)$ and $c(x, \mu = 0.4, \boldsymbol{\beta = 14}, \Delta W^{env} = -0.3, \Delta W^{non} = 0.3, \rho = 1)$. In both cases, it is clear that for low x, the majority of the bi-logistic curve owes its shape to the green logistic curves, which correspond to the choices of the *environmentalist* individuals. When x is larger, the effect of the *non−environmentalist* individuals becomes more apparent. The dashed green and brown lines are the Fisher-Pry transforms of the corresponding logistic functions. Notice that the Fisher-Pry transforms are parallel; they both have the same gradient, which is $a = 2\beta\rho$.

In the $\beta = 7$ case, the green (*environmentalist*) curve is still significantly increasing at the point on the x axis where the brown curve begins to become significant. This is reflected in the behaviour of the Fisher-Pry transforms: the



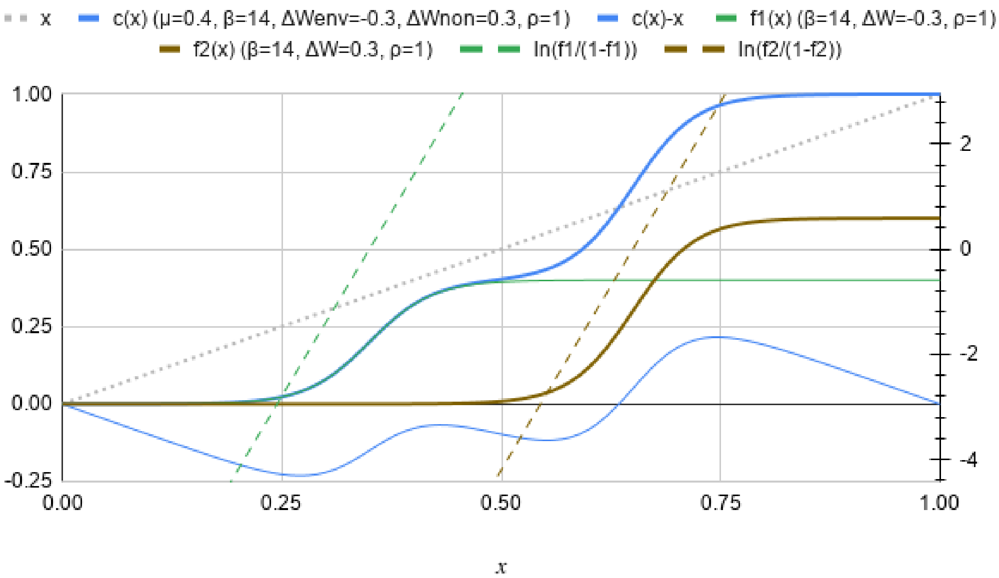


Figure 9: Fisher-Pry transforms of two bi-logistic curves. The green solid lines are the logistic curves corresponding to the *environmentalist* individuals, and the brown solid lines are the logistic curves for the $non-environmentalist$ individuals. Their dashed counterparts are the Fisher-Pry transforms of these logistic functions. The upper graph has $\beta = 7$ and the lower graph $\beta = 14$. Also shown are the $y = x$ line (in grey), and $c(x) - x$ (in light blue), whose roots are the points of equilibria in each case. The right hand axis is for the Fisher-Pry transforms (only) and has the scale adjusted so that 1 on the left-hand axis corresponds to 95% on the right-hand axis, 0.5 corresponds to 0 on the right-hand axis, and 0 to 5%.

two dashed lines overlap somewhat, as opposed to the behaviour in the $\beta = 14$ example below.

Meyer (1994) provides a useful taxonomy of the summation of logistic curves. The $\beta = 7$ graph is an example of "superposed" logistic functions, and the $\beta = 14$ graph exemplifies the "sequential" type. The other two categories are "converging" and "diverging," both of which have lines with dissimilar gradients, characterising logistic curves with different values for a, and therefore need not concern us here.

3.4 Dynamics

We will show:

- That $0 \leq x_I^{env} \leq ½ \leq x_I^{non} \leq 1$;
- That there is at least one and at most five equilibria;
- If there are five equilibria x_1, x_2, x_3, x_4, x_5 with $x_1 < x_2 < x_3 < x_4 < x_5$, then $x_1 < \frac{1}{2}$ and $x_5 > \frac{1}{2}$;
- If there are five equilibria then x_1, x_3 and x_5 are stable and x_2 and x_4 are unstable;
- If there is one equilibrium then it is stable;
- If $\mu = 0.5$ and $\Delta W^{env} = -\Delta W^{non}$ then the curve has rotational symmetry around the point $(\frac{1}{2}, \frac{1}{2})$

As before, we are also able to make statements concerning the conditions required for the existence of five equilibria.

Proposition 3.1

$0 \leq x_I^{env} \leq ½ \leq x_I^{non} \leq 1.$

Proof. Result follows from $x_I^t = b = \frac{1}{2} + \frac{\Delta W^t}{2\rho}$ and $\Delta W^{env} \leq \frac{1}{2} \leq \Delta W^{non}$ with equality only when $\Delta W^{env} = \Delta W^{non} = 0$. □

Proposition 3.2

If $\Delta W^{env} = \Delta W^{non}$, then the heterogeneity disappears and the solution is as in section 2.

Proof.

$$\Delta W^{env} = \Delta W^{non} \Rightarrow \; f(x, \beta, \Delta W^{env}, \rho) = f(x, \beta, \Delta W^{non}, \rho)$$

So

$$\begin{aligned} c(\mu, x, \beta, \Delta W^{non}, \Delta W^{non}, \rho) &= \mu f(x, \beta, \Delta W^{env}, \rho) + (1 - \mu) f(x, \beta, \Delta W^{env}, \rho) \\ &= f(x). \end{aligned}$$

□

Note that this applies even if the base utility received by *environmentalist* and *non − environmentalist* individuals is different; all that is required is for the differences ΔW^t to be equal.

This makes sense intuitively—if the ΔW^t are equal, then all individuals face the same marginal cost decision, no matter which type they belong to. The result will therefore be as in a homogeneous discrete choice experiment.

Proposition 3.3

There is at least one equilibrium.

Proof. $c(0) > 0$ and $c(1) < 1$, and $c(x)$ is everywhere continuous and differentiable. □

Proposition 3.4

There are at most five equilibria.

Proof. The first derivative of a logistic function is a single-humped function, with a single maximum. The summation of the first derivative of two logistic functions can therefore have no more than two maxima. The largest number of times any straight line can intersect such a curve is five. □

Proposition 3.5

If there are five equilibria x_1, x_2, x_3, x_4, x_5 with $x_1 < x_2 < x_3 < x_4 < x_5$, then $x_1 < \frac{1}{2}$ and $x_5 > \frac{1}{2}$;

Proof. If $c(x)$ has five equilibria $\in [0, 1]$ then $c(x) - x = 0$ has five roots, which means the derivative, $\frac{dc}{dx} - 1$ changes sign four times. We know from proposition 3.1 that $x_I^{env} \leq \frac{1}{2} \leq x_I^{non}$. We cannot have equality because then we would have $\Delta W^{env} = \Delta W^{non} = 0$, and proposition 3.2 would apply, limiting the number of equilibria to three. So $x_I^{env} < \frac{1}{2} < x_I^{non}$. Now suppose that there are no equilibria $\in [0, \frac{1}{2})$. Then there must be five in the range $(\frac{1}{2}, 1]$. But $x_I^{env} < \frac{1}{2}$, so $c(x) - x$ can change signs at most four times which means only four roots—a contradiction. Therefore there must be at least one equilibrium in the range $[0, \frac{1}{2})$. If there are more than one, then we choose the smallest, and name it x_1. By a similar process, there exists at least one equilibrium point $x_5 > \frac{1}{2}$. □

Proposition 3.6

If there are five equilibria then x_1, x_3 and x_5 are stable equilibria and x_2 and x_4 are unstable;

Proof. If there are five equilibria, then the sign of $c(x) - x$ alternates between positive and negative, starting as a positive (because $c(0) > 0$) and ending with a negative ($c(1) < 1$). The sign changes exactly at the five points of equilibria. At the first, third and fifth of these, the sign of $c(x) - x$ goes from positive to negative, so $\frac{dc}{dx} < 1$ at each point. This is precisely the requirement for a stable equilibrium. □

Proposition 3.7

If there is one equilibrium then it is stable;

Proof. (This is essentially the same proof as in the section 2 case, reprised here for ease of reference.) $c(0) > 0$ and $c(1) < 1$, so there is an $x_1 \in (0, 1)$ such that $c(x_1) - x = 0$. For values of x near this point, $g(x) > 0$ for $x < x_1$ and $g(x) > 0$ for $x > x_1$. Thus successive iterations of x_t will converge towards this equilibrium, making it stable. □

Proposition 3.8

If $\mu = 0.5$ and $\Delta W^{env} = -\Delta W^{non}$ then the curve has rotational symmetry around the point $(\frac{1}{2}, \frac{1}{2})$ (see figure 10).

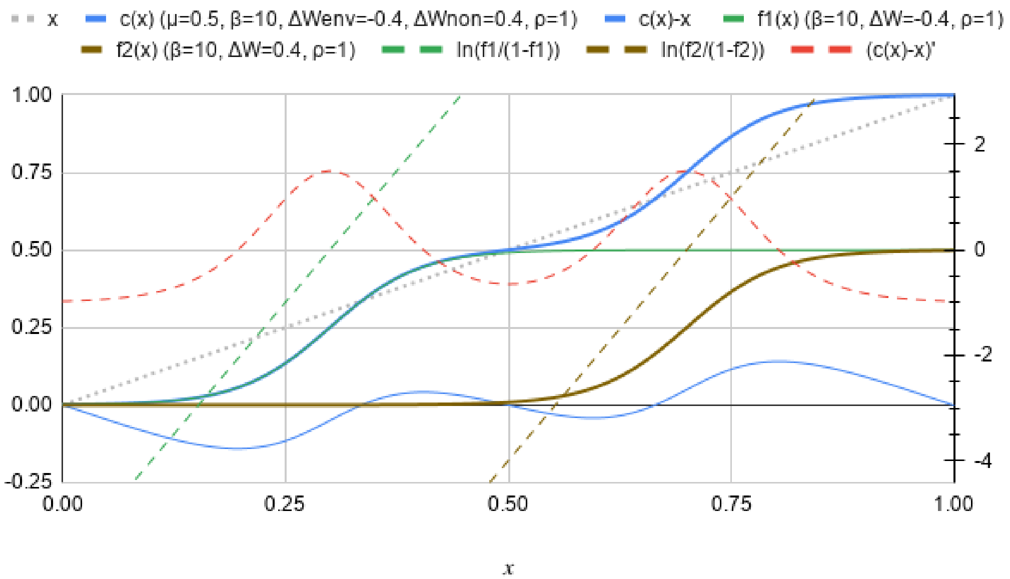


Figure 10: An example of the symmetric situation arising in proposition 3.8. The curve passes through $(\frac{1}{2}, \frac{1}{2})$, about which point it is rotationally symmetric.

Proof. Let $z = x - b$, and write $c(x) = \mu f_1(x) + (1-\mu) f_2(x)$. Then

$$1 - c(-z) = 1 - \big(\mu f_1(-z) + (1-\mu) f_2(-z)\big).$$

But $f_i(-z) = 1 - f_i(z)$ by proposition 2.6, so this equals

$$\begin{aligned}
&= 1 - \big(\mu(1 - f_1(z)) + (1-\mu)(1 - f_2(z))\big) \\
&= 1 - \mu + \mu f_1(z) - 1 + \mu + f_2(z) - \mu f_2(z) \\
&= \mu f_1(z) + (1-\mu) f_2(z) \\
&= c(z)
\end{aligned}$$

$\square$

3.5 Classification

Counting glancing equilibria, there are as many as 17 different classes of solutions to the model in this section[2].

There are however, only five different numbers of equilibria, and a maximum of only three different "outcomes" (stable equilibria) which is where the interest lies.

Recall that in section 2, if an indifference point existed then there were two different outcomes: individuals converged either on the *sustainable* or the *unsustainable* option, depending on the strength of parameters.

In this model, there is now an additional outcome: the population can converge on a "hybrid" solution, in which part of the population makes one choice while the rest makes the other. This new outcome is only possible when there are five equilibria.

3.6 Summary

Provided that the thresholds exist, then there are now *three* potential outcomes. Recall that in the base model, the only possible outcomes were that everyone chose *unsustainable* or that everyone chose *sustainable*. Now there is a third

[2] For completeness, we have included a summary of each of the 17 types in the Appendix (starting on page 43).

outcome, which pertains whenever there are five equilibria in total, and can be found when $x_3 \leq x \leq x_4$.

This intermediate outcome is interesting. It is stable, which means that the overall number of people choosing the *sustainable* option is static, but there is a disagreement between the *environmentalist* and the $non-environmentalist$ populations as to what to choose.

The intuition behind this outcome is straightforward; it corresponds to a situation in which most of the *environmentalist* individuals choose the *sustainable* option and the $non - environmentalist$ individuals choose the *unsustainable* option. If this seems unexciting, that is probably because it is so commonplace in real world scenarios. It is however encouraging to have a model which is capable of sustaining this kind of a tension between agents; that is a feature not always found in mathematical models.

4 Conclusion

We have seen that adding additional agent types to this kind of discrete choice model can give rise to outcomes (in the form of stable equilibria) which were not previously admissible.

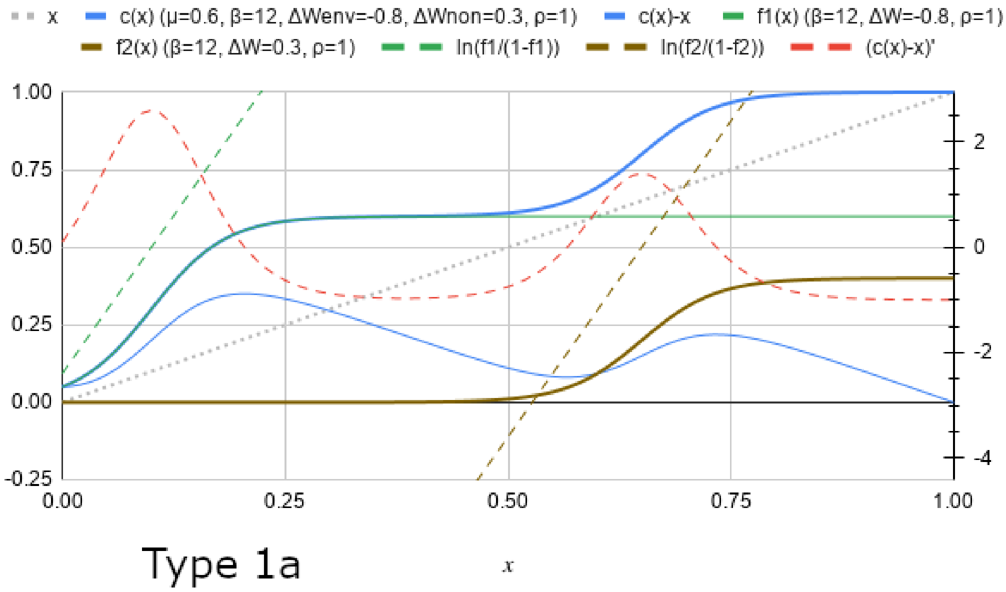
It is easy see how to extend this model to multiple (> 2) agent types. The resulting distribution could be given as:

$$x_t = Pr\{\omega_i = k\} = \sum \frac{\mu_k}{1 + e^{\beta(\Delta W^k + \rho(1-2x_{t-1}))}} \tag{18}$$

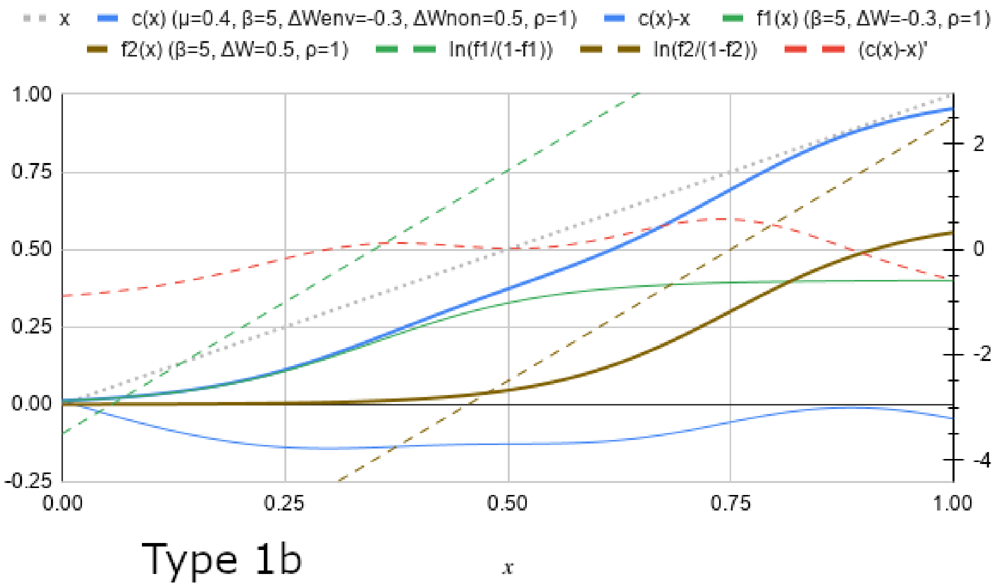
the graph of which would in general resemble a series of sigmoid "humps," of varying sizes (given by μ_k) and positions (given by ΔW^k).

While it would not be possible to calculate exact values with this model, it is fortunately highly computationally tractable, and the Fisher-Pry transforms make it eminently suitable for empirical modelling.

Appendix A Bi-logistic Classification



Type 1a



Type 1b

Figure 11: Single equilibrium types. Above, type 1a, below, type 1b. The dashed lines use the right-hand axis, all others use the left-hand axis. Both these cases have a single stable equilibrium, and are distinguished purely by whether their equilibria are closer to 0 or 1.

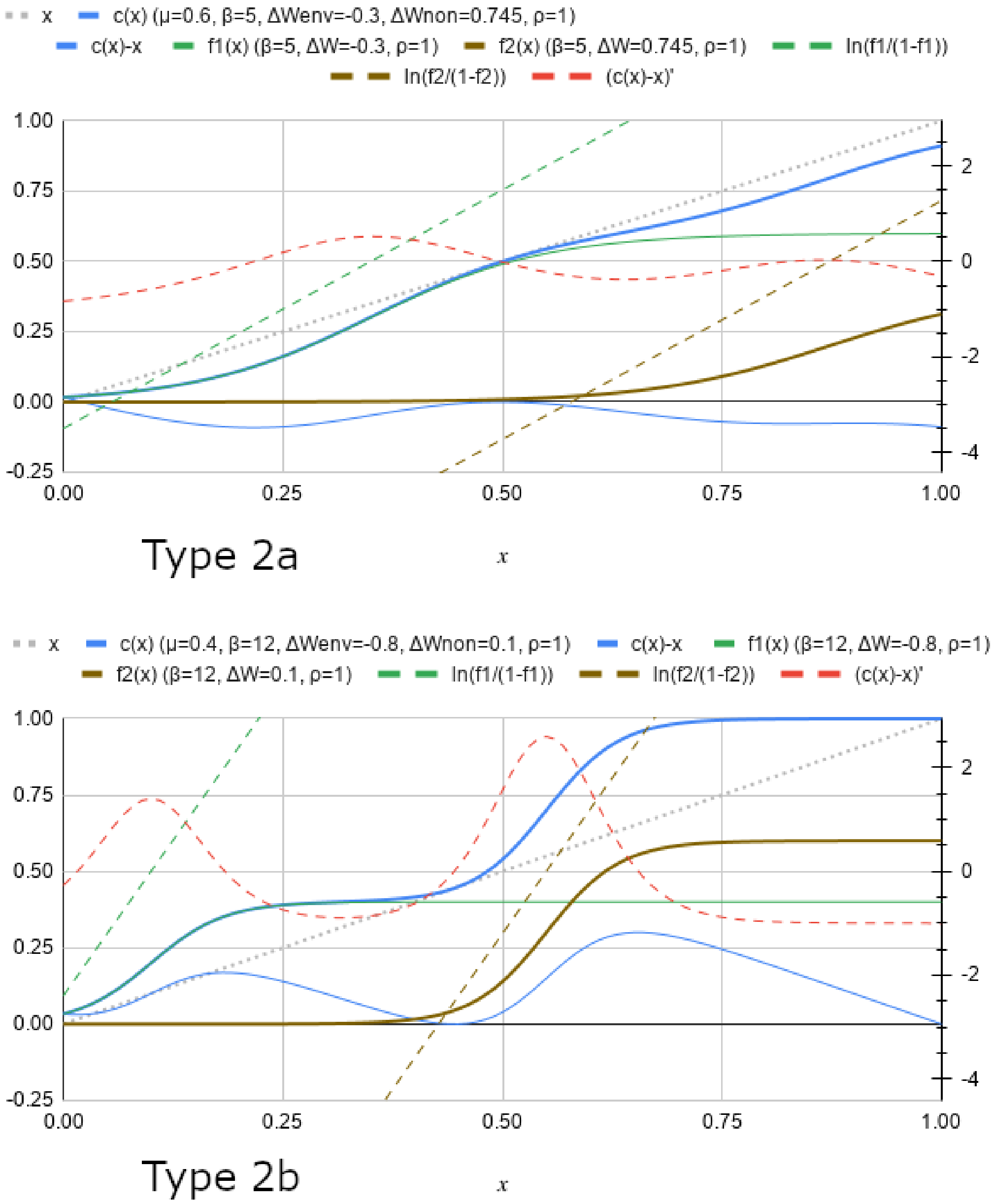


Figure 12: Two equilibria, types 2a (above) and 2b. These both have a single stable equilibrium and a single (glancing) unstable equilibrium. The example of type 2a happens also to be a clear example of "sequential" Fisher-Pry transforms, indicating that there is scope for both logistic curves to be "in operation" concurrently.

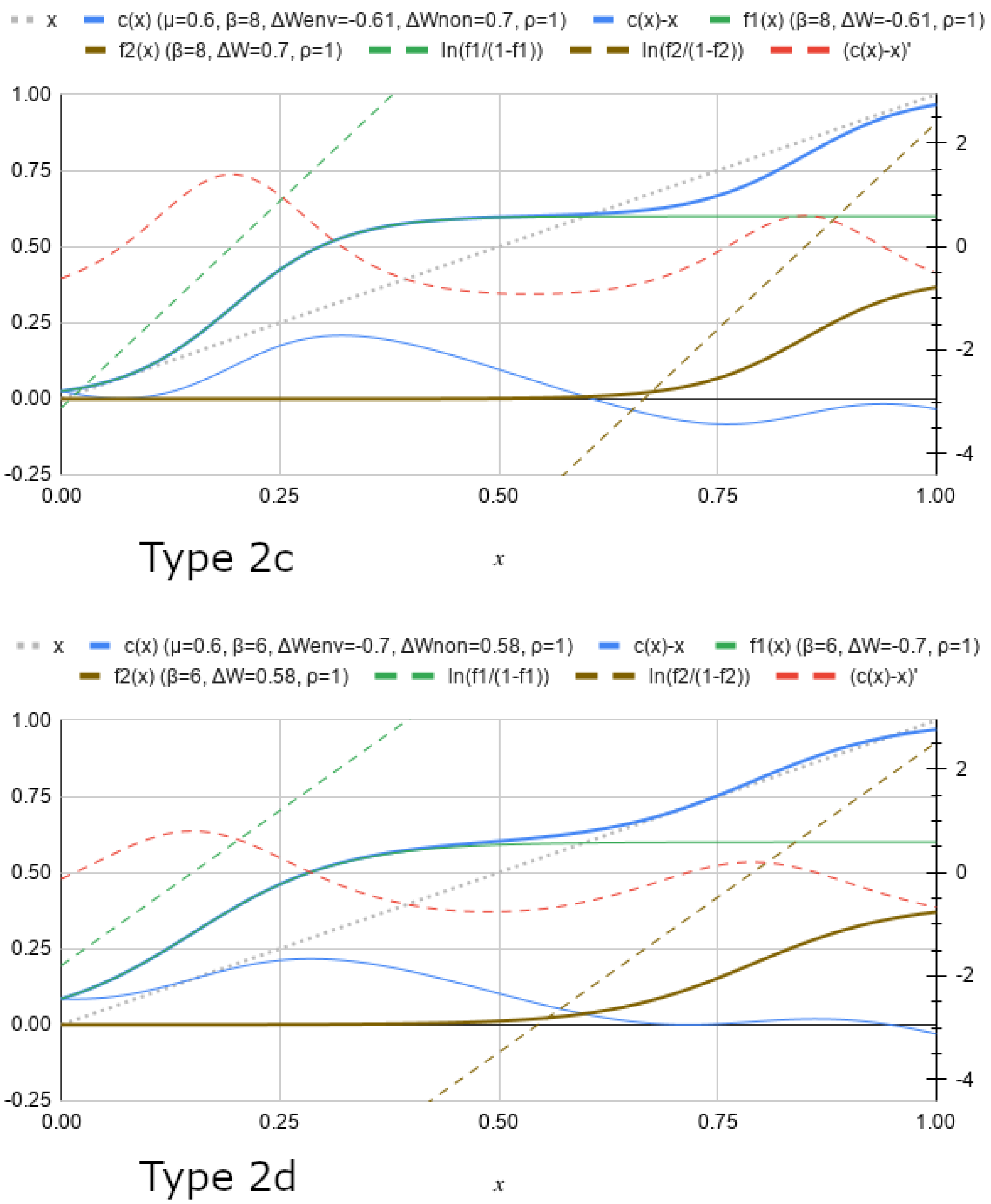


Figure 13: Two equilibria, types 2c (above) and 2d. As with figure 12, both types have a single unstable equilibrium and a single (glancing) unstable one.

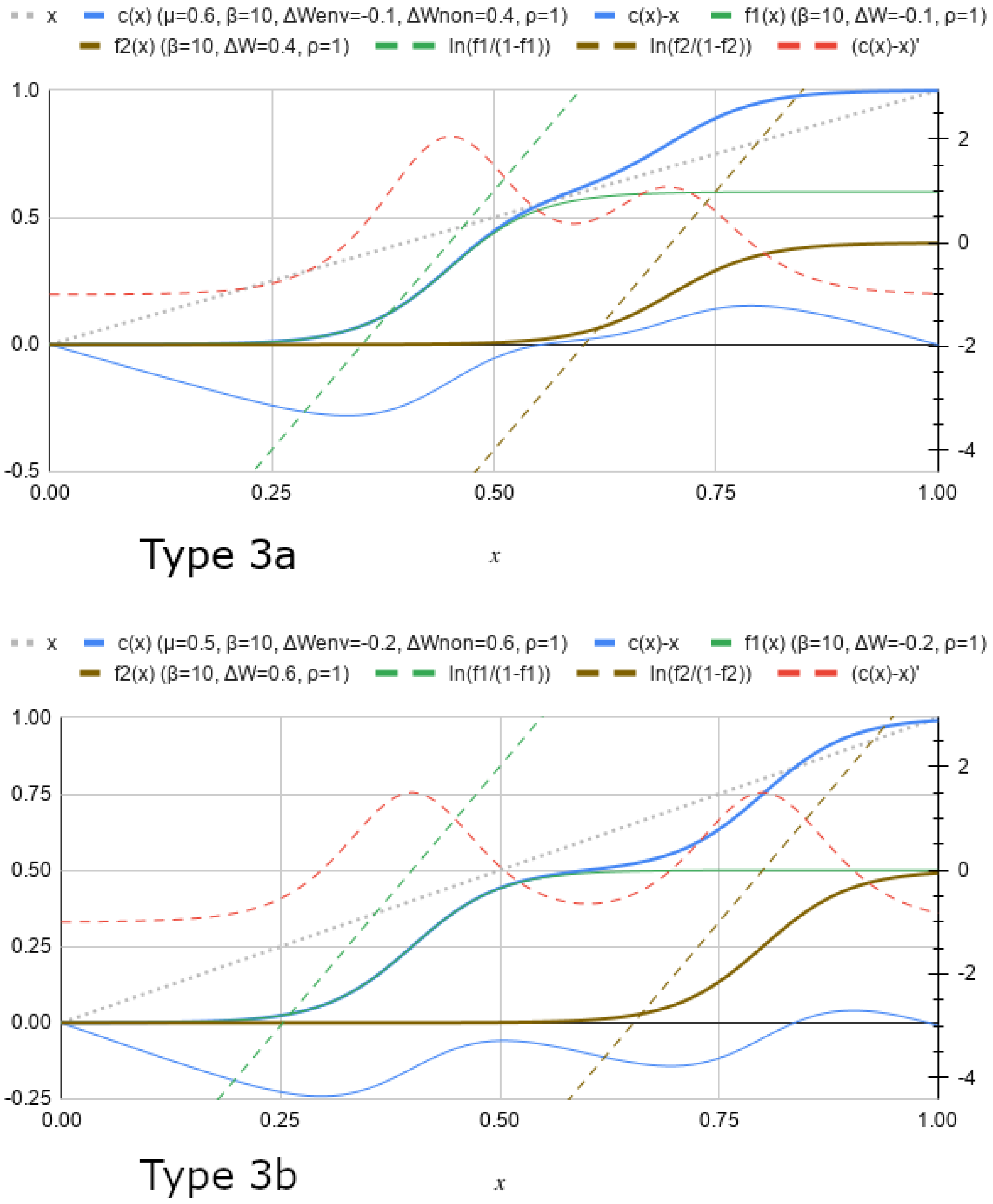


Figure 14: Three equilibria, types 3a (above) and 3b. Type 3a is the type which is most similar to the three-equilibrium case for logistic curves. For both types, and as with the logistic curves, the middle equilibrium is stable and the two extreme equilibria are unstable.

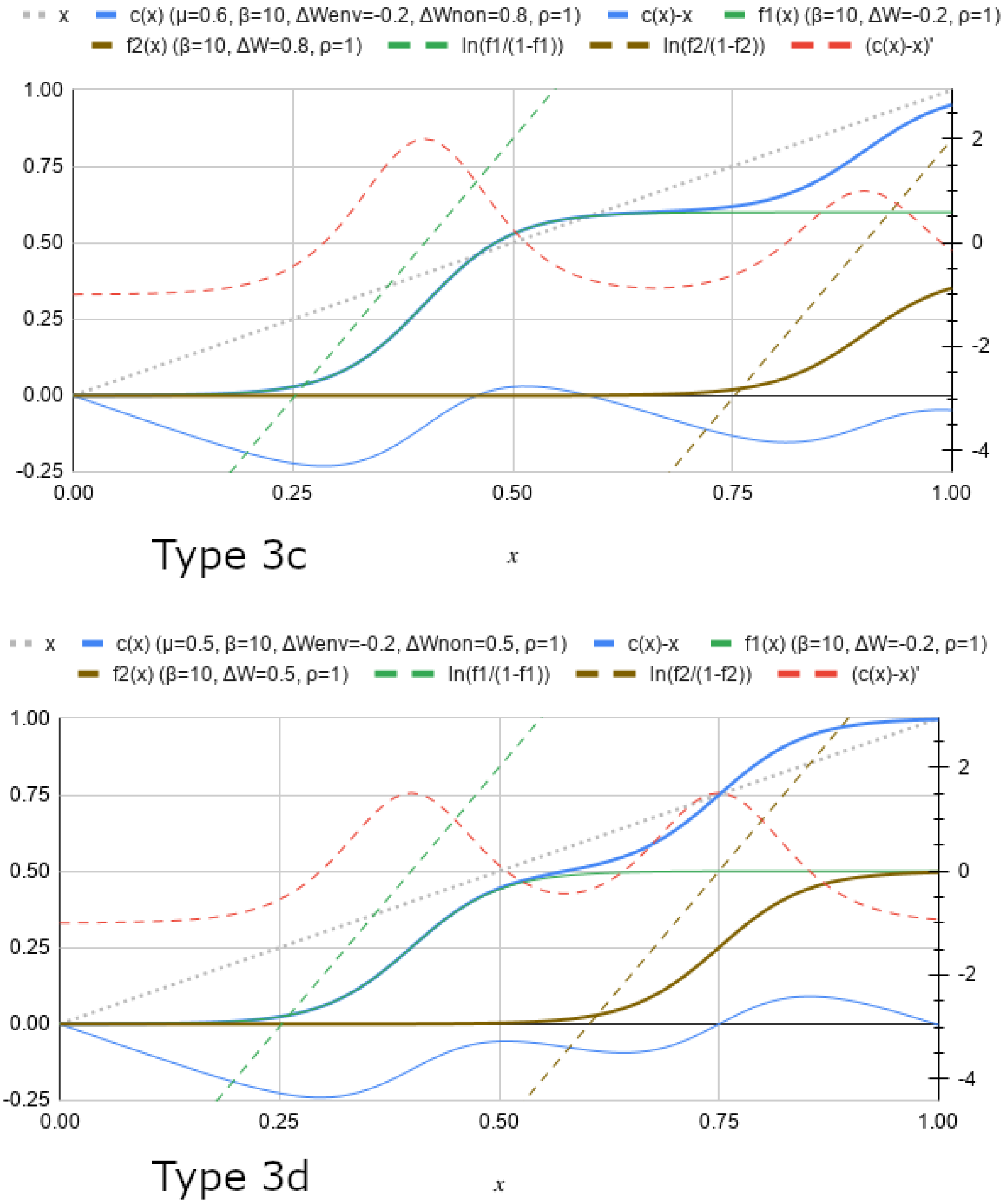
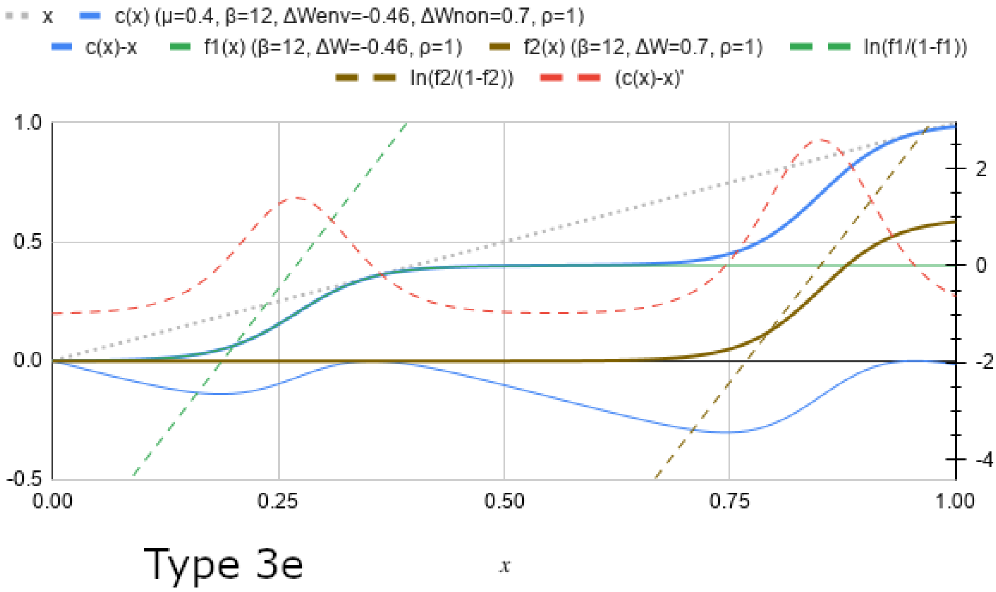
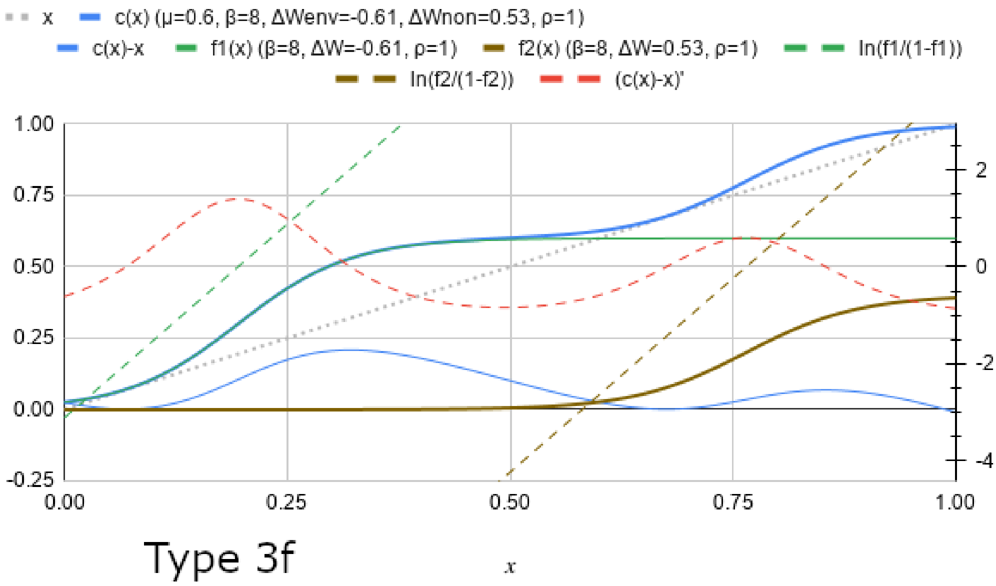


Figure 15: Three equilibria, types 3c (above) and 3d. As with figure 14, the middle equilibrium is stable and the two extreme equilibria are unstable.

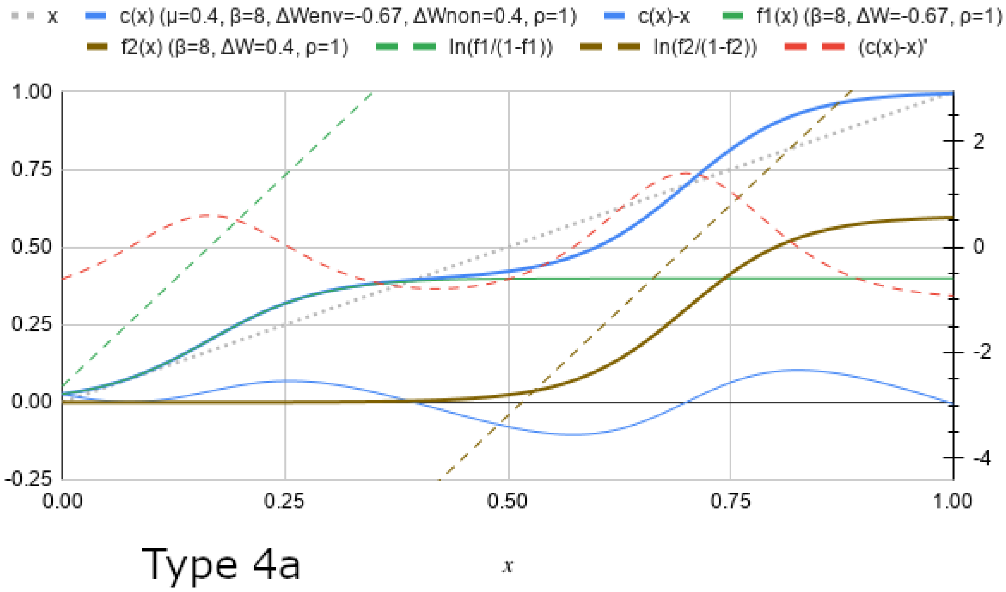


Type 3e



Type 3f

Figure 16: Three equilibria, types 3e (above) and 3f. Both types are examples of two glancing (semi-stable) equilibria; type 3e has the equilibria as local maxima, 3f as local minima. For both types, iterations converge towards the stable equilibrium; for type 3e, this is x_1, for type 3f, x_3.



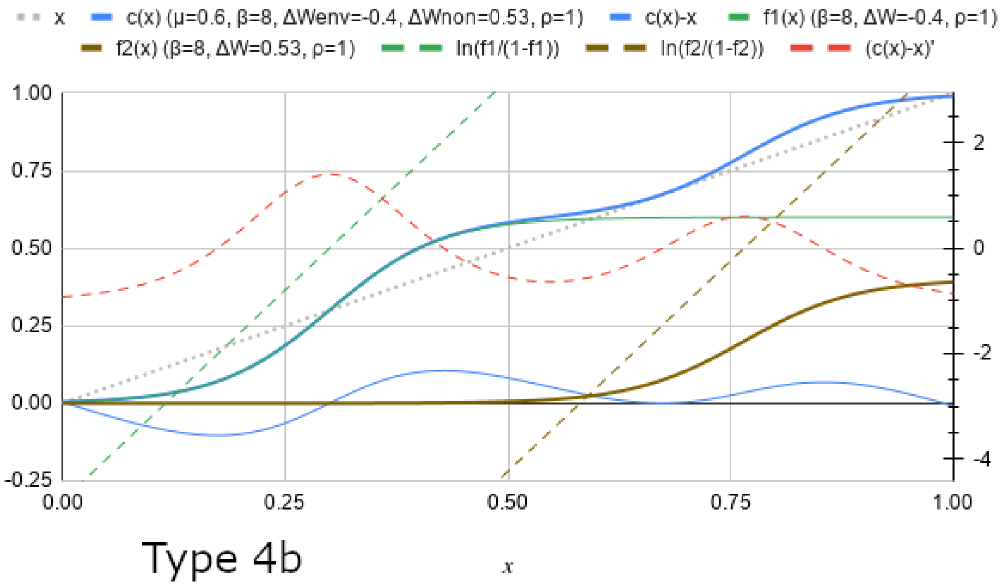
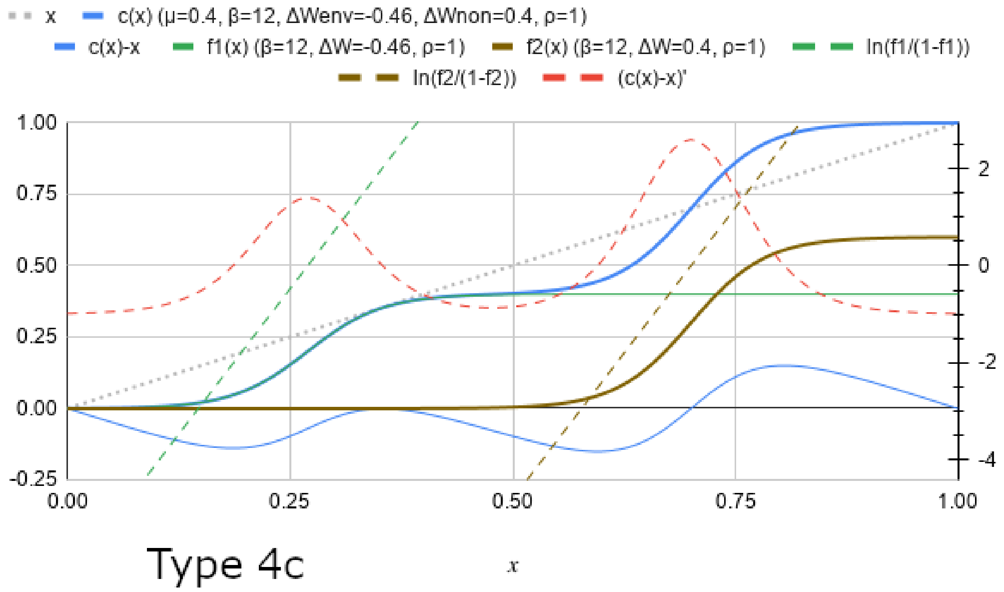


Figure 17: Four equilibria, types 4a (above) and 4b. These cases, as with types 4c and 4d in figure 18, have one glancing (semi-stable) equilibrium and three other equilibria. In all cases, the lower and upper of the non-glancing equilibria are the stable ones, and the other one is unstable.



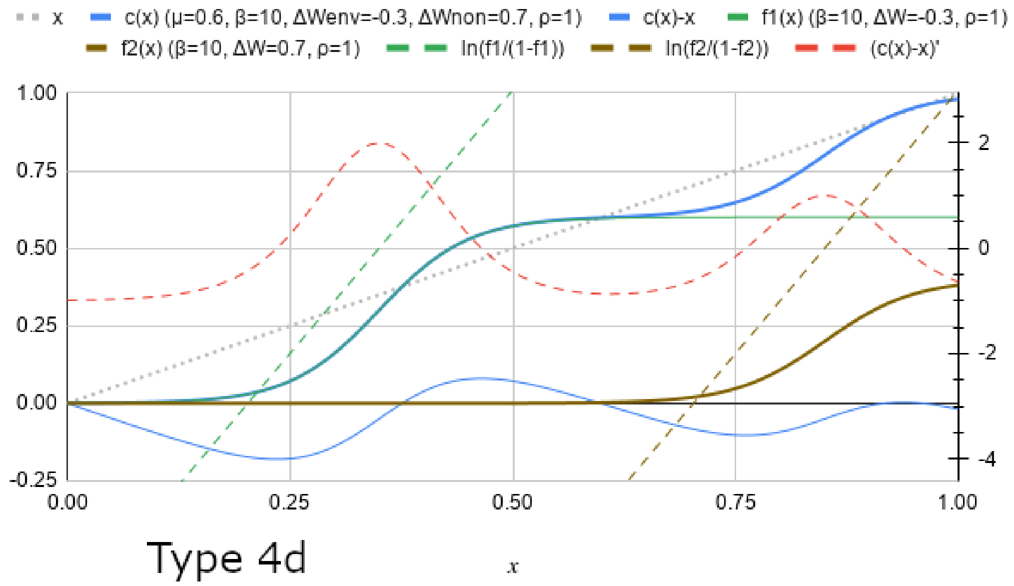


Figure 18: Four equilibria, types 4c (above) and 4d. These cases, as with types 4a and 4b in figure 18, have one glancing (semi-stable) equilibrium and three other equilibria. In all cases, the lower and upper of the non-glancing equilibria are the stable ones, and the other one is unstable.

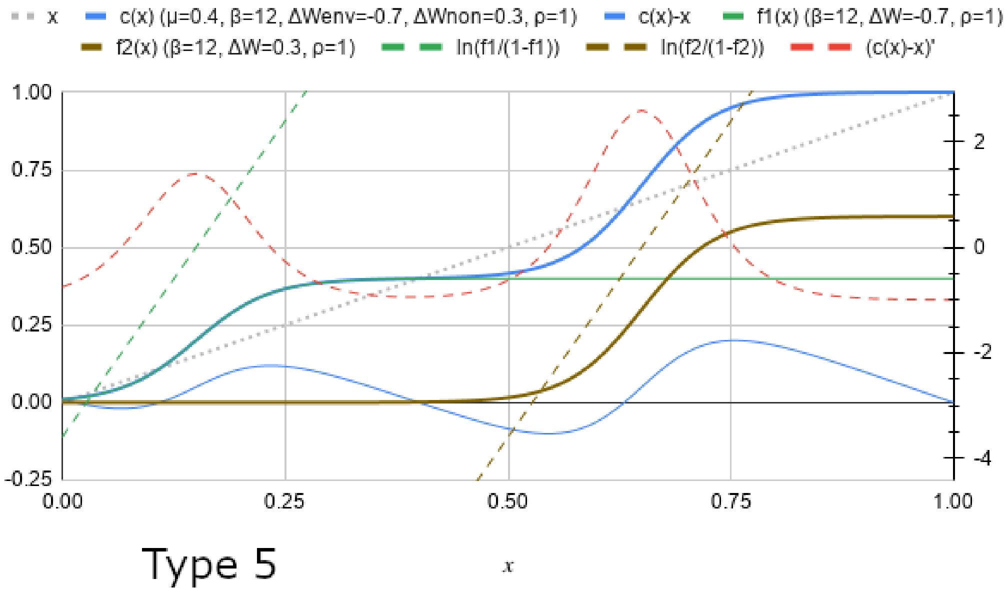


Figure 19: Five equilibria (type 5, above). x_1, x_3 and x_5 are stable. This is the only one of the 17 types which supports three stable equilibria.

References

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L. and Maguire, K.C., 2011. Has the Earths sixth mass extinction already arrived? *Nature*, 471(7336), p.51.

Bicchieri, C., 2006. *The grammar of society*. Cambridge University Press.

Brock, W.A. and Durlauf, S.N., 2001. Discrete choice with social interactions. *Review of economic studies*, 68(2), pp.235–260.

den Elzen, M., Kuramochi, T., Höhne, N., Cantzler, J., Esmeijer, K., Fekete, H., Fransen, T., Keramidas, K., Roelfsema, M. and Sha, F., 2019. Are the G20 economies making enough progress to meet their NDC targets? *Energy policy*, 126, pp.238–250.

Feldmann, J. and Levermann, A., 2015. Collapse of the West Antarctic Ice Sheet after local destabilization of the Amundsen Basin. *Proceedings of the national academy of sciences*, 112(46), pp.14191–14196.

Fisher, J.C. and Pry, R.H., 1971. A simple substitution model of technological change. *Technological forecasting and social change*, 3, pp.75–88.

Foster, G., Royer, D. and Lunt, D., 2017. Future climate forcing potentially without precedent in the last 420 million years. *Nature communications*, 8, p.14845.

Ge, M., Lebling, K., Levin, K. and Friedrich, J., 2019. *Tracking progress of the 2020 climate turning point.* (Technical report). World Resources Institute.

Hamilton, C., 2015. Why we resist the truth about climate change. *Iii congreso internacional sobre cambio climático y desarrollo sustentable (la plata, 2011)*.

Hamilton, L.C., Hartter, J., Lemcke-Stampone, M., Moore, D.W. and Safford, T.G., 2015. Tracking public beliefs about anthropogenic climate change. *Plos one*, 10(9), e0138208.

Hasselmann, K., Latif, M., Hooss, G., Azar, C., Edenhofer, O., Jaeger, C.C., Johannessen, O.M., Kemfert, C., Welp, M. and Wokaun, A., 2003. The challenge of long-term climate change. *Science*, 302(5652), pp.1923–1925.

Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Ballew, M., Goldberg, M. and Gustafson, A., 2018. *Climate change in the American mind: December 2018*. (Technical report). Yale University and George Mason University, New Haven, CT.

Manski, C.F. and McFadden, D., 1981. *Structural analysis of discrete data with econometric applications*. mit Press Cambridge, MA.

Marshall, G., 2015. *Don't even think about it: why our brains are wired to ignore climate change*. Bloomsbury Publishing USA.

Masson-Delmotte, V., Zhai, P., Pörtner, H.O., Roberts, D., Skea, J., Shukla, P.R. and Waterfield, T., 2018. *Global warming of 1.5 degrees C*. (Technical report). The Intergovernmental Panel on Climate Change.

Meyer, P., 1994. Bi-logistic growth. *Technological forecasting and social change*, 47(1), pp.89–102.

Newton, I., 1687. *Philosophiae naturalis principia mathematica; auctore isaaco newtono, eq. aurato; perpetuis commentariis illustarta, communi studio pp. thomae le seur & francisci jacquier, ex gallicana minimorum familia, matheseos professorum: tomus primus*. Vol. 1.

Prieur, J., 2020. Critical warning! preventing the multidimensional apocalypse on planet Earth. *Ecosystem services*, 45, p.101161.

Thøgersen, J., 2008. Social norms and cooperation in real-life social dilemmas. *Journal of economic psychology*, 29(4), pp.458–472.

Tong, D., Zhang, Q., Zheng, Y., Caldeira, K., Shearer, C., Hong, C., Qin, Y. and Davis, S.J., 2019. Committed emissions from existing energy infrastructure jeopardize 1.5 degrees C climate target. *Nature*, 572(7769), pp.373–377.

UNFCCC, 2015. *Paris Agreement*. (Technical report). UNFCCC.

Viscusi, W.K., Huber, J. and Bell, J., 2011. Promoting recycling: private values, social norms, and economic incentives. *American economic review*, 101(3), pp.65–70.

Zeppini, P., 2015. A discrete choice model of transitions to sustainable technologies. *Journal of economic behavior & organization*, 112, pp.187–203.