Chalotorn Kansuntisukmongkol
ERG, UC-Berkeley, 03/31/04

Irreversibility and Learning in Global Climate Change Problem: Linear and Single-Quadratic Models

1. Introduction

The concept of "irreversibility effect" is originally proposed by Arrow and Fisher (1974) and Henry (1974). They have demonstrated that, for a binary-choice or linear utility model, if there is uncertainty about the costs/benefits of the action choices and if one of the binary choices is irreversible, a decision maker would find it beneficial to err her decision away from the irreversible choice when there is a possibility of learning about the uncertainty in the future compared to the case when there is no future learning. The "irreversibility effect" is regarded to be a distinct, though related, feature from risk-aversion.

Subsequent literature has tried to extend the concept of "irreversibility effect" to non binary-choice, non-linear model. It is found that the "irreversibility effect" always holds in the intertemporally separable model, but may not hold universally when the model is intertemporally non-separable. A number of sufficient and necessary conditions have been proposed to identify the criteria when the "irreversibility effect" would hold (Epstein 1980; Freixas and Laffont 1984; Ulph and Ulph 1995; Kolstad 1996; Ulph and Ulph 1997; Gollier et al. 2000; Narain et al. 2003)

This paper applies a modified version of Epstein's Theorem to a very simplified model of global climate change problem. The results show that the conventional conception of the "irreversibility effect" does not always hold even in the most simplistic cases (linear and single-quadratic models). This indicates a fundamental flaw in the "irreversible effect" framework.

This paper argues that the key to better understanding of the issue is to recognize that the current definition of "irreversibility effect" does not necessarily have anything to do with irreversibility per se (Epstein 1980; Ulph and Ulph 1995; Ulph and Ulph 1997). The focus of the current definition of "irreversibility effect" is on the intertemporal effects of future learning, not the effects of irreversibility itself. Moreover, the intertemporal characteristic of the "irreversibility effect" also implies that we need to take both types of intertemporal dependencies – i.e., constraint-dependency and function-dependency – into consideration in our analysis.

This paper proposes a new framework in addressing the question based on two new concepts: "learning effect" and "irreversibility bias". The paper illustrates how the "learning effect when no-constraint" and the "irreversibility bias on the learning effect" work in creating the "learning effect / irreversibility effect" in the irreversible case. Utilizing a special characteristic of the linear and single-quadratic models, a simple approach that can be used to readily identify the direction of "irreversibility bias" in these models is proposed. The results show that the direction of "irreversibility bias", as well as the direction of "learning effect / irreversibility effect", is much more dependent on the functional assumptions of the model than currently expected in the literature. This indicates the importance of the functional dependency mechanism in influencing the direction of "learning effect / irreversibility effect".

The rest of the paper is organized as follows. Section 2 describes a general model and the standard definition of “irreversibility effect”. Section 3 and 4 explain the simplified model of global climate change problem and the methodology in analyzing the “irreversibility / learning effect”. Section 5 discusses the results of the analysis from the “irreversibility effect” perspective as well as the inconsistencies found in the results. The new framework is proposed in Section 6. Detailed analysis of two single-quadratic models is presented in Section 7 to illustrate how the new framework operates. Section 8 discusses a simple approach (Simple Analysis 1) that can be used to identify the direction of the “irreversibility bias” in linear and single-quadratic cases. Section 9 provides a conclusion.

2. General Model and the Definition of "Irreversibility Effect"

Consider the following general two-period optimization problem (following Epstein, 1980):

(1) $$J = \underset{x_1 \in C_1}{Max}\left\{E_Y\left[\underset{x_2 \in C_2(x_1)}{Max}\left\{E_{Z/Y}[U(x_1, x_2, Z)]\right\}\right]\right\}$$

where x_1 and x_2 are control variables for the two periods. C_1 and $C_2(x_1)$ are convex subset of the real number line representing the permissible choices of x_1 and x_2, respectively. Z is a random variable representing the uncertainty in the utility function. The true value of Z becomes known at the end of period two. The decision maker, however, gains some knowledge about Z through the observation of another random variable Y (an experiment) at the beginning of period two, before she makes her decision on x_2.

Y and Z are assumed to be discrete random variables with possible realizations $(y_1,\ldots,y_n)$ and $(z_1,\ldots,z_m)$ respectively. Let q and r denote the decision maker's subjective probability vectors of Y and Z, i.e., $q_j = Pr(Y=y_j)$ and $r_i = Pr(Z=z_i)$. After learning about Y, the planner updates her expectation about Z following Bayes' Rule. Π denotes the posterior probability distribution after the planner learns about Y ($\pi_{ij} = Pr(z=z_i| y=y_j)$ and $\pi_j = (\pi_{1j},\ldots,\pi_{mj})$).

For simplicity, consider only the following two extreme experiments:

1. Full Learning (L): This is the case when Y = Z and the learning of Y would reveal all the information of Z. Specifically, this means that n = m, $\pi_{ij} = 1$ when i = j and $\pi_{ij} = 0$ otherwise.
2. No Learning (NL): This is the case when Y and Z are perfectly independent and the learning of Y reveals no information of Z. Specifically, this means that $\pi_{ij} = r_i$ for all j.

Also consider two possible sets of constraints on x_1 and x_2:

1. No-constraint Case (NC): $C_1 = \{x_1 \in \Re\}$ and $C_2 = \{x_2 \in \Re\}$
2. Irreversible Case (IR): $C_1 = \{x_1 \in \Re\}$ and $C_2 = \{x_2 \in \Re \mid x_2 \geq x_1\}$.

Let $x_1^{L\text{-}NC}$, $x_1^{NL\text{-}NC}$, $x_1^{L\text{-}IR}$, and $x_1^{NL\text{-}IR}$ denotes the optimal solutions of x_1 for the full-learning & no-constraint, no-learning & no-constraint, full-learning & irreversible, and no-learning & irreversible cases, respectively. According to the literature, the "irreversibility effect" can be defined as follows:

Definition 1. The irreversibility effect holds, if and only if $C_2(x_1^{L\text{-}IR}) \supseteq C_2(x_1^{NL\text{-}IR})$. In the case of our model, this is equivalent to the condition that $x_1^{L\text{-}IR} \leq x_1^{NL\text{-}IR}$.

3. A Simple Model of Global Climate Change Problem

To facilitate the discussion, this study uses a very simplified model of global climate change problem as the case study. Consider the following two-period decision problem in which the social planner tries to choose the optimal levels of greenhouse gases (GHG) emissions, x_1 for period one and x_2 for period two, that would minimize the sum of GHG abatement costs and GHG-induced damage costs over the two periods. Assuming that the discount factor is one (i.e., there is no discounting), the objective function can be imprecisely written in general form[1] as:

$$(2) \quad J = \underset{0 \leq x_1 \leq X_1}{Min} \left\{ O_1(A_1) + I_1(A_1 - A_0) + D_1(G_1, \theta) + \underset{f_1(x_1) \leq x_2 \leq f_2(x_1)}{Min} \left\{ O_2(A_2) + I_2(A_2 - A_1) + D_2(G_2, \theta) \right\} \right\}$$

with x_t = GHG emissions in period t
X_t = potential GHG emissions in period t that would occur if there is no abatement
A_t = GHG abatement level in period t = $X_t - x_t$
G_t = stock of atmospheric GHG in period t
$f_1(x_1)$ = minimum allowable choices of x_2 (which may depend on the choice of x_1)
$f_2(x_1)$ = maximum allowable choices of x_2 (which may depend on the choice of x_1)
θ = random variable denoting uncertainty in the model

The abatement cost is assumed to consist of two components. The operating and maintenance cost ($O_t(A_t)$) represents all the variable costs in operating and maintaining the abatement facilities. It is assumed to be a direct function of the abatement level in each period (A_t). The investment cost ($I_t(A_t - A_{t-1})$) represents the costs of expanding the abatement capacity. It is assumed to be a function of the change in abatement level compared to the previous period ($A_t - A_{t-1}$). The damage cost ($D_t(G_t,\theta)$) is assumed to be a function of the stock of GHGs in the atmosphere (G_t) and a random variable (θ) which reflects the underlying uncertainty in the model. The model assumes that the only uncertain component in the model is the damage cost function, thus the random variable θ does not appear in the abatement cost functions. The three cost functions are assumed to be convex and twice continuously differentiable in A_t, $A_t - A_{t-1}$, and G_t, respectively.

Similar to the general model, this model assumes that when the social planner makes her decision in period one, she is uncertain about the size of the random variable θ but has some prior information regarding its probability distribution. For simplicity, it is assumed that $\theta \geq 0$ and θ appears in the damage cost function in a linear multiplicative manner – i.e., $D_1(G_1, \theta) = \theta \cdot \widetilde{D}_1(G_1)$. The damage $D_1(\cdot)$ occurs at the end of period one and provides a chance for the planner to learn about the true value of θ. To simplify the analysis, the model considers only the two extreme possibilities of learning – i.e., no learning (NL) and full learning (L). The true value of θ is known after $D_2(\cdot)$ occurs at the end of period two.

[1] The precise form depends on further assumption concerning the future learning of θ which will be discussed below.

To further simplify the analysis, the following additional assumptions are assumed:

(1) The stock of atmospheric GHG in period 0 (before period 1) is zero ($G_0 = 0$ and thus $G_1 = x_1$).
(2) The stock of GHG abatement capital and, thus, the level of GHG abatement in period 0 are zero ($A_0 = 0$).
(3) The potential GHG emissions that would occur if there is no abatement for both period 1 and 2 are the same ($X_1 = X_2 = X$).
(4) The stock of atmospheric GHG decays naturally at the rate of 1-Δ. Thus the stock of atmospheric GHG at the end of period 2 becomes $G_2 = x_2+\Delta x_1$.
(5) The level of GHG abatement is a linear function of the stock of abatement capital and the stock of capital depreciates at the rate of 1-δ. Thus under the assumption of capital irreversibility, the minimum limit of GHG abatement would be that $A_2 \geq \delta A_1$.

Under the above assumptions, the objective function when there is full learning at the end of period one becomes:

$$\underset{x_1}{Min}\left\{E_\theta\left[O_1(X-x_1)+I_1(X-x_1)+D_1(x_1,\theta)+\underset{x_2(x_1,\theta)}{Min}\left\{O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,\theta)\right\}\right]\right\}$$

$$(3)\qquad =\underset{x_1}{Min}\left\{O_1(X-x_1)+I_1(X-x_1)+D_1(x_1,E\theta)+E_\theta\left[\underset{x_2(x_1,\theta)}{Min}\left\{O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,\theta)\right\}\right]\right\}$$

Under the above assumptions, the objective function when there is no learning at the end of period one becomes:

$$\underset{x_1}{Min}\left\{E_\theta\left[O_1(X-x_1)+I_1(X-x_1)+D_1(x_1,\theta)+\underset{x_2(x_1)}{Min}\left\{E_\theta[O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,\theta)]\right\}\right]\right\}$$

$$(4)\qquad =\underset{x_1}{Min}\left\{O_1(X-x_1)+I_1(X-x_1)+D_1(x_1,E\theta)+\underset{x_2(x_1)}{Min}\left\{O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,E\theta)\right\}\right\}$$

The analysis is conducted in four different models: a linear model and three single-quadratic models. In these models, the three cost functions (O(·), I(·), and D(·)) are assumed to be either linear or quadratic. For the linear model, all cost functions are assumed to be linear. For the single-quadratic models, only one of the three cost functions is assumed to be quadratic while the other two are still linear.[2]

The specific functional forms for each of the three cost functions are as follows[3]:

Operating and maintenance cost

O-linear: $O_t(\cdot) = \alpha_t A_t = \alpha_t(X - x_t)$

O-quad: $O_t(\cdot) = \alpha_{t1}A_t + \alpha_{t2}A_t^2 = \alpha_{t1}(X - x_t) + \alpha_{t2}(X - x_t)^2$

[2] The analysis and discussions for double- and triple-quadratic models will be presented in my next paper (Essay2).

[3] All parameters, α, β, and γ, are assumed to be non-negative. To keep the model simple, the constant terms are omitted from the cost functions. But adding the constant terms would not affect the results of the analysis. Please note that all quadratic cost functions are convex with respect to their corresponding variables.

Investment cost

I-linear: $I_t(\cdot) = \beta_t(A_t - A_{t-1}) \quad = \beta_t(x_{t-1} - x_t)$

I-quad: $I_t(\cdot) = \beta_{t1}(A_t - A_{t-1}) + \beta_{t2}(A_t - A_{t-1})^2 \quad = \beta_{t1}(x_{t-1} - x_t) + \beta_{t2}(x_{t-1} - x_t)^2$

Damage cost

D-linear: $D_t(\cdot) = \theta\gamma_t G_t \quad = \theta\gamma_t(x_t + \Delta G_{t-1})$

D-quad: $D_t(\cdot) = \theta\gamma_{t1}G_t + \theta\gamma_{t2}G_t^{\ 2} \quad = \theta\gamma_{t1}(x_t + \Delta G_{t-1}) + \theta\gamma_{t2}(x_t + \Delta G_{t-1})^2$

Table 1: List of Models

Model	**Functional Assumptions**			**Appropriate Choice of Control Variable**[4]
1	D-linear	O-linear	I-linear	All variables (including x_t)
2	**D-quad**	O-linear	I-linear	G_t
3	D-linear	**O-quad**	I-linear	x_t
4	D-linear	O-linear	**I-quad**	Q_t

4. Methodology of Analysis

The methodology of analysis used in this paper is based on the theorem proposed by Epstein (1980) and further clarified by Ulph and Ulph (1997). Considering the above two-period control problem, we can rewrite the objective functions in the full learning and no-learning cases as:

$$\begin{aligned}
J(L) &= \underset{x_1}{Min}\ E\,[O_1(X-x_1)+I_1(X-x_1)+D_1(x_1,\theta)+ \underset{x_2(x_1,\theta)}{Min}\,[O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,\theta)]] \\
&= \underset{x_1}{Min}\ E\,[\,V(x_1,\theta) + \underset{x_2(x_1,\theta)}{Min}\,[O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,\theta)]] \\
&= \underset{x_1}{Min}\ E\,[\,V(x_1,\theta) + C(x_1,\theta)]
\end{aligned}$$

$$\begin{aligned}
J(NL) &= \underset{x_1}{Min}\ E\,[O_1(X-x_1)+I_1(X-x_1)+D_1(x_1,\theta)+ \underset{x_2(x_1)}{Min}\ E\,[O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,\theta)]] \\
&= \underset{x_1}{Min}\ E\,[\,V(x_1,\theta) + \underset{x_2(x_1)}{Min}\,[O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,E\,\theta)]] \\
&= \underset{x_1}{Min}\ E\,[\,V(x_1,\theta) + C(x_1,E\,\theta)]
\end{aligned}$$

with $V(x_1,\theta) = O_1(X-x_1)+I_1(X-x_1)+D_1(x_1,\theta)$ and $C(x_1,\theta) = \underset{x_2(x_1,\theta)}{Min}\,[O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1,\theta)]$ for learning case and $C(x_1,E\,\theta) = \underset{x_2(x_1)}{Min}\,[O_2(X-x_2)+I_2(x_1-x_2)+D_2(x_2+\Delta x_1, E\theta)]$ for no-learning case. A modified/simple version of Epstein's theorem can be stated as follows.

[4] See the discussion in Section 8.

<u>Theorem 1</u>: *Assuming that $V(x_1,\theta)$ and $C(x_1,\theta)$ are convex in x_1*

(a) if $C_{x_1}(x_1,\theta)$ is convex in θ, then $x_1^L \leq x_1^{NL}$

(b) if $C_{x_1}(x_1,\theta)$ is concave in θ, then $x_1^L \geq x_1^{NL}$

(c) if $C_{x_1}(x_1,\theta)$ is linear in θ, then $x_1^L = x_1^{NL}$.

<u>Proof:</u> The first order conditions characterizing the optimal values of x_1 for learning and no learning cases are:

Learning: $$V_{x1}(x^L, E\theta) + \sum_{\theta} \pi_\theta C_{x1}(x^L,\theta) = 0$$

No Learning: $$V_{x_1}(x^L, E\theta) + C_{x_1}(x^L, E\theta) = 0$$

If the marginal cost function $C_{x_1}(x_1,\theta)$ is concave (convex) in θ, then the expected marginal cost at any particular value of x_1 is lower (higher) under learning than under no-learning. By assuming that both $V(x_1,\theta)$ and $C(x_1,\theta)$ are convex, the attainment of both of the first order conditions would require that we have $x_1^L \geq x_1^{NL}$ ($x_1^L \leq x_1^{NL}$). If $C_{x_1}(x_1,\theta)$ is linear in θ, it is both (weakly) concave and (weakly) convex in θ. Thus, we must have $x_1^L = x_1^{NL}$.

Under standard economic assumptions, we would expect the function $V(\cdot)$ to be convex in x_1. Moreover, if we have the constraint set of x_2 that is "well-behaved", we could also expect to have function $C(\cdot)$ that is convex in x_1 (Epstein, 1980). (An example of a "well-behaved" constraint set is when both constraints $f_1(x_1)$ and $f_2(x_1)$ are linear.) Since both irreversibility constraints discussed in this paper are linear, the only remaining question that we need to consider in determining the direction of "irreversibility effect" in this model is whether $C_{x_1}(\cdot)$ is a concave or convex function in θ. Thus, the discussion on the results of each model in the following sections is conducted based on the graph of $C_{x_1}(\cdot)$ on the θ axis.

5. Results from the "Irreversibility Effect" Perspective

This section describes the results of the analysis of the "irreversibility effect" following the methodology explained in Section 4. Four different stock irreversibility/reversibility constraints are considered in the analysis. These constraints include: GHG stock irreversibility (G-irr; $G_2 \geq \Delta G_1$ or $x_2 \geq 0$), GHG stock reversibility (G-rev; $G_2 \geq 0$ or $x_2 \geq -\Delta x_1$), abatement capital stock irreversibility (K-irr; $A_2 \geq \delta A_1$ or $x_2 \leq X-\delta(X-x_1)$)[5], and abatement capital reversibility (K-rev; $A_2 \geq 0$ or $x_2 \leq X$).

[5] To make the K-irr assumption comparable to the existing literature, it is interpreted as a constraint on the minimum amount of GHG abatement such that all the inherited abatement capital (minus depreciation) must always be utilized. Our separation of GHG abatement cost into operating and maintenance cost (O_t) and investment cost (I_t), however, indicates an inconsistency in the existing models that consider the K-irr assumption. For such K-irr constraint to actually exist, all the GHG abatement costs must be in the form of capital cost and there must be no variable cost component. Most models, especially the numerical models, however, are constructed as if the GHG abatement costs consist of only the variable cost component.

Please note that the reversible assumptions (G-rev and K-rev) used here are not the same as the "no constraint" assumption implicitly assumed in all existing literature when the "no irreversibility" case is considered.[6]

The standard convention of "irreversibility effect" expects that the G-irr will induce an "irreversibility effect" that biases towards less current emissions ($x_1^L \leq x_1^{NL}$) while the K-irr will create another "irreversibility effect" that biases towards more emissions ($x_1^L \geq x_1^{NL}$). Since none of the existing literature has considered the G-rev and K-rev constraints, I would only say that the conventional conception is violated when the G-rev constraint has induced an "irreversibility effect" that biases towards more emissions or when the K-rev constraint has created an "irreversibility effect" that biases towards less emissions. Please note that the conventional conception expects the "no-irreversibility" (no-constraint) assumption to create no "irreversibility effect"[7].

Table 2: Analytical Results on the Direction of "Irreversibility Effect"

Constraint	**Model 1 (app var = x_t)**	**Model 2 (app var = G_t)**	**Model 3 (app var = x_t)**	**Model 4 (app var = Q_t)**
K-irr	Higher x_1	Higher x_1	Higher x_1	*Lower x_1*
K-rev	Neutral	Higher x_1	Neutral	*Lower x_1*
G-irr	*Neutral*	Lower x_1	*Neutral*	*Higher x_1*
G-rev	*Higher x_1*	Neutral	*Higher x_1*	*Higher x_1*

Table 2 shows the summary results on the direction of the "irreversibility effect" for each irr/rev constraint in each model (see Appendix C for details of the analysis). Half of the results (presented in italic) are found to be inconsistent with the conventional conception regarding "irreversibility effect". The K-irr constraint does not always induce the "irreversibility effect" that biases towards more emissions and the G-irr constraints does not always induce the "irreversibility effect" that biases towards less emissions. The K-rev and G-rev constraints are also found to induce the wrong direction of "irreversibility effect" as well. Moreover, the direction of the "irreversibility effect" of each constraint is found to be highly dependent on the functional assumptions of the model. The most interesting results occur in Model 4 – where the K-irr and K-rev constraints induce the "irreversibility effect" that biases towards less emissions while the G-irr and G-rev constraints induce the "irreversibility effect" that biases towards more emissions. In other words, all the constraints are found to induce the "irreversibility effects" in the wrong directions compared to the conventional conception. These results indicate an important flaw in the conventional conception of the "irreversibility effect" and perhaps the need for a new framework.

[6] Ulph and Ulph (1995, 1997) explicitly discuss the G-rev assumption but use the "no constraint" assumption in their actual analysis.
[7] For all linear or single-quadratic models, the no-constraint assumption will always create no "irreversibility effect" since the marginal cost function C_{x1} when there is no constraint in these models is always linear in θ (see Theorem 1). This result, however, does not always hold for more complex models.

6. A New Framework

Epstein (1980) is the first to propose a general criterion that could be used to determine whether the "irreversibility effect" holds or not. It is important, however, to note that Epstein does not frame his problem as a study of the effects of irreversibility, but as "an investigation of the effects on decision making of the temporal resolution of uncertainty" (Epstein 1980, p.270). This is indeed the correct description of the problem that economists focus on when we analyze the "irreversibility effect".

Consider Definition 1, the focus of the analysis of "irreversibility effect" is whether $x_1^{L\text{-}IR}$ is larger or smaller than $x_1^{NL\text{-}IR}$. We are comparing the two optimal solutions of x_1 that differ "only with respect to the amount of information provided by Y about Z" (Epstein 1980, p.269). The analysis of "irreversibility effect" is not a study of the effects of irreversibility – since we are not comparing the optimal solutions of the irreversible and no-constraint cases. The two optimal solutions we are comparing assume exactly the same irreversibility constraint but differ only on the amount of future learning. Thus, the analysis of "irreversibility effect" is actually the investigation of the effects of future learning on the optimal decision under the context of uncertainty and irreversibility. In the current definition of "irreversibility effect", irreversibility is just a part of the context of analysis not the focus of the analysis.

The incorrect framing of the problem causes at least two types of unresolved anomalies.

- The "irreversibility effect" may not hold in the irreversible case – i.e., $x_1^{L\text{-}IR} > x_1^{NL\text{-}IR}$. All the inconsistencies found in Section 5 fall into this type of anomalies.
- The "irreversibility effect" could hold even if there is no "irreversibility constraint" – i.e., $x_1^{L\text{-}NC} \leq x_1^{NL\text{-}NC}$. See my next paper (Essay 2) for examples of this type of anomalies.

To resolve these anomalies[8], I propose two new concepts: "learning effect" and "irreversibility bias on the learning effect". The *"learning effect"* concerns the comparison between the optimal solution when there is no-learning and the solution when there is full learning. Thus, it is essentially equivalent to the conventional concept of "irreversibility effect", but is defined more generally.

Given a particular context M (i.e., a set of assumptions M about the situation).

> *__Definition 2.1__ The learning effect pushes towards higher x_1, if $x_1^{L\text{-}M} > x_1^{NL\text{-}M}$,*
> *__Definition 2.2__ The learning effect pushes towards lower x_1, if $x_1^{L\text{-}M} < x_1^{NL\text{-}M}$,*
> *__Definition 2.3__ There is no learning effect, if $x_1^{L\text{-}M} = x_1^{NL\text{-}M}$.*

The concept of "learning effect" is more general than the definition of "irreversibility effect" in two senses. (i) There is no predetermined direction that the learning effect is supposed to hold. The direction of learning effect is described independently from the irreversible/reversible constraint.

[8] Narain et al. (2003) propose a different definition for the "irreversibility effect" – i.e., the "irreversibility effect" holds if $\left|x_1^{L-IR} - \hat{x}_1\right| \leq \left|x_1^{NL-IR} - \hat{x}_1\right|$ with $\hat{x}_1$ being the choice of x_1 that gives maximum decision making flexibility in the future. This alternative definition could alleviate some of the anomalies discussed here, but it would not eliminate all of them – since the problem is still being framed as the study of the effect of irreversibility and not the effect of future learning.

(ii) The concept of learning effect is equally applicable for the irreversible, reversible, or no-constraint cases. Thus, it is possible to have the learning effect that pushes towards one direction in the no-constraint case and pushes towards the other direction in the irreversible case.

The second concept, *"irreversibility bias on the learning effect"*, addresses how the addition of irreversibility constraint into the model changes the direction of the learning effect compared to the no-constraint case. Thus, the direction of the "learning effect with irreversibility constraint" is determined by both the "learning effect when no-constraint" and the "irreversibility bias on the learning effect". The precise definition of "irreversibility bias on the learning effect" will be described in the section when we examine the details of the analysis on the "learning effect".

The most important advantage of the use of these new concepts over the "irreversibility effect" definition is that it brings us to the correct framing of the problem under consideration. Framing the problem as a study of intertemporal interaction between learning and decision making processes allows us to understand the problem better. Since "learning effect" is an intertemporal feature, it can only occur when there is an intertemporal connection between our current-period decision and our future decisions. If there is no intertemporal connection – i.e., if what we do today does not in any way affect our future decisions – then we can consider them as separate decisions to be made independently. This implies that whatever is going to happen in the future – including whether or not we are going to have future learning about uncertainty – becomes irrelevant to our current decision. This means there is no "learning effect". It is only when there is an intertemporal connection among decisions at different time periods that we will have "learning effect".

There are two possible types of intertemporal connection: *intertemporal constraint-dependency* and *intertemporal functional-dependency*. Intertemporal constraint-dependency occurs when our choice in current period has an effect in limiting or enhancing the range of available options in our future period decisions – this has been regarded as the focal point of the studies on “irreversibility effect”. However, there is another type of intertemporal dependency that could happen. Intertemporal functional-dependency refers to the case when our current period choice does not directly affect the range or availability of future period options, but it affects the marginal cost and/or marginal benefit function of the future options. The change in marginal cost/benefit function then affects the optimal choices that we choose in the future decisions. Therefore, similar to constraint-dependency, functional-dependency also requires that we take this induced change in the future optimal path into account in selecting our current optimal choice. The fact that the “irreversibility effect” does not hold universally when the benefit/cost functions are intertemporally non-separable indicates the importance of the functional-dependency. Any analysis of "learning / irreversibility effect" that does not take the functional assumptions of the model into account could not be considered complete.

Thus, in order to address the importance of both dependency mechanisms, the analysis in this paper is conducted on a group of models with different functional and constraint assumptions.

7. Results from the "Learning Effect" Perspective

This section illustrates how the proposed framework operates through the detailed analysis of two of the four models (i.e., Model 2 and 4). These two models are selected because of their contrasting results. As shown in Table 2, Model 2 yields results that are fully consistent with the conventional conception while Model 4 gives results that totally contradict the conventional conception of "irreversibility effect". Thus the analysis of these two models should provide a sufficient coverage of the new framework. Moreover, in order to allow a more complete explanation, the analysis in this section will be limited to only the two irreversible constraints (G-irr and K-irr). The analysis for the other two constraints and/or the other two models can be conducted in an analogous manner.

It is important to note that since the concepts of "learning effect" and the "irreversibility effect" are essentially equivalent, the same analytical method (as described in Section 4) is applied here. The major difference between the two frameworks is how they interpret the results.

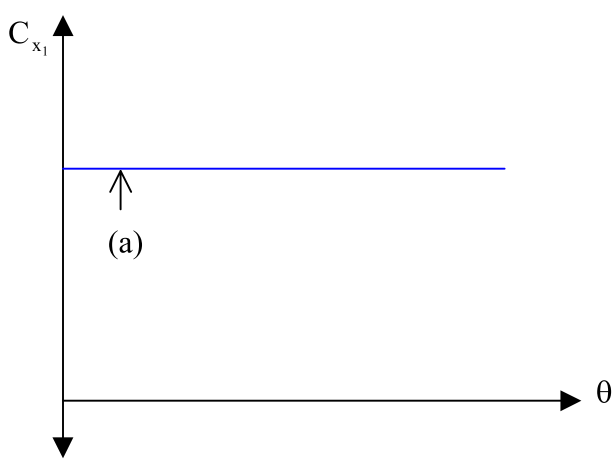
<u>Analysis for Model 2</u> (D-quad, I- and O-linear)

Model 2 assumes that $D_t(\cdot)$ is a quadratic function while $O_t(\cdot)$ and $I_t(\cdot)$ are linear functions. Under such assumptions, the objective functions under full learning and no-learning become:

$$
\begin{aligned}
J_L &= \underset{x_1}{Min}\,[\alpha_1(X-x_1)+\beta_1(X-x_1)+E\,\theta\gamma_{11}x_1+E\,\theta\gamma_{12}x_1{}^2+E\,[\underset{x_2(x_1,\theta)}{Min}\,[\alpha_2(X-x_2)+\beta_2(x_1-x_2)\\
&\qquad +\theta\gamma_{21}(x_2+\Delta x_1)+\theta\gamma_{22}(x_2+\Delta x_1)^2]]]\\
&= \underset{x_1}{Min}\,[V^2(x_1, E\,\theta)+E\,C^2(x_1,\theta)].
\end{aligned}
$$

$$
\begin{aligned}
J_{NL} &= \underset{x_1}{Min}\,[\alpha_1(X-x_1)+\beta_1(X-x_1)+E\,\theta\gamma_{11}x_1+E\,\theta\gamma_{12}x_1{}^2+\underset{x_2(x_1)}{Min}\,[\alpha_2(X-x_2)+\beta_2(x_1-x_2)+E\,\theta\gamma_{21}(x_2+\Delta x_1)\\
&\qquad +E\,\theta\gamma_{22}(x_2+\Delta x_1)^2]]\\
&= \underset{x_1}{Min}\,[V^2(x_1, E\,\theta)+C^2(x_1, E\,\theta)].
\end{aligned}
$$

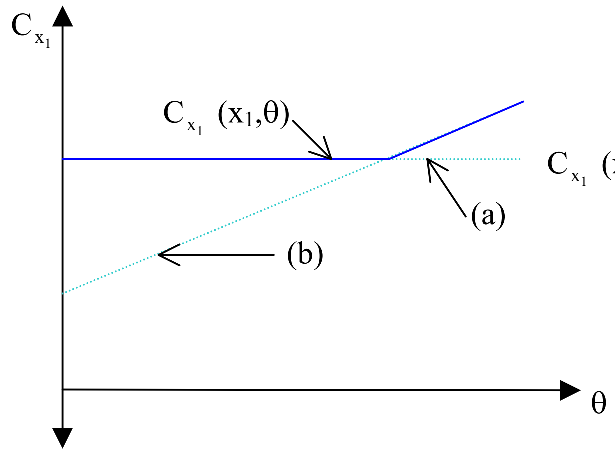
Figure 1a, 1b, 1c, and 1d show the graphs of $C^2_{x_1}(\cdot)$ on the θ axis for Model 2 when there is no-constraint, when there is only G-irr constraint, when there is only K-irr constraint, and when there are both G-irr and K-irr constraints, respectively.



$C_{x_1}(x_1,\theta) = \beta_2+\alpha_2\Delta+\beta_2\Delta = (a)$

Figure 1a: No Constraint Case in Model 2

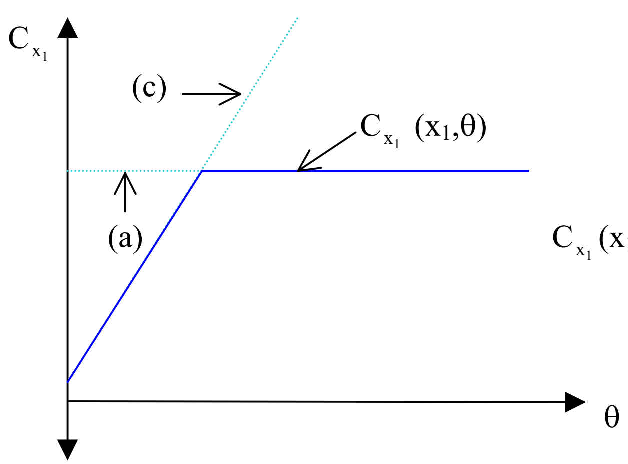
Consider Figure 1a. When there is no irreversibility constraint in Model 2, $C_{x_1}(\cdot)$ is found to be a linear function in θ – i.e., $C_{x_1}(x_1,\theta) = \beta_2+\theta\gamma_{21}\Delta+2\theta\gamma_{22}\Delta(x_2^*+\Delta x_1) = \beta_2+\alpha_2\Delta+\beta_2\Delta$ = (a). Thus, Model 2 has no learning effect when there is no constraint.



$$C_{x_1}(x_1,\theta) = \begin{cases} \beta_2+\alpha_2\Delta+\beta_2\Delta = (a) \\ \qquad \text{when } \theta \le (\alpha_2+\beta_2)/(\gamma_{21}+2\gamma_{22}\Delta x_1) \\ \beta_2+\theta\gamma_{21}\Delta+2\theta\gamma_{22}\Delta^2 x_1 = (b) \\ \qquad \text{when } \theta > (\alpha_2+\beta_2)/(\gamma_{21}+2\gamma_{22}\Delta x_1) \end{cases}$$

Figure 1b: G-irr Case in Model 2

Consider Figure 1b for Model 2 when there is G-irr constraint only. When G-irr constraint is inactive or when $\theta \le (\alpha_2+\beta_2)/(\gamma_{21}+2\gamma_{22}\Delta x_1)$, we have the interior solution as the optimal solution and thus $C_{x_1}(x_1,\theta)$ = (a). When G-irr constraint is active or when $\theta > (\alpha_2+\beta_2)/(\gamma_{21}+2\gamma_{22}\Delta x_1)$, the optimal solution is to choose $x_2 = 0$ (following the G-irr constraint) and we have $C_{x_1}(x_1,\theta) = \beta_2+\theta\gamma_{21}\Delta+ 2\theta\gamma_{22}\Delta^2 x_1$ = (b). Both (a) and (b) are linear functions in θ, but when we couple them together, the coupling of these two segments of $C_{x_1}(\cdot)$ becomes a convex function and we have $C_{x_1}(\cdot)$ = max [(a), (b)]. Since $C_{x_1}(\cdot)$ is convex in θ, there is a "learning effect" in this case that pushes towards lower x_1 (i.e., $x_1^L \le x_1^{NL}$).



$$C_{x_1}(x_1,\theta) = \begin{cases} -\alpha_2\delta + \beta_2(1-\delta) + \theta\gamma_{21}(\delta+\Delta) \\ + 2\theta\gamma_{22}(\delta+\Delta)(X-\delta X+\delta x_1+\Delta x_1) = (c) \\ \qquad \text{when } \theta < \dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}(X-\delta X+\delta x_1+\Delta x_1)} \\ \beta_2+\alpha_2\Delta+\beta_2\Delta = (a) \\ \qquad \text{when } \theta \ge \dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}(X-\delta X+\delta x_1+\Delta x_1)} \end{cases}$$

Figure 1c: K-irr Case in Model 2

Figure 1c shows the graph of $C_{x_1}(x_1,\theta)$ for Model 2 with K-irr constraint. When the K-irr constraint is inactive or when $\theta \geq \dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}(X-\delta X+\delta x_1+\Delta x_1)}$, the optimal solution is the interior solution and we have $C_{x_1}(x_1,\theta)$ = (a). When K-irr constraint is active or when $\theta < \dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}(X-\delta X+\delta x_1+\Delta x_1)}$, we have $C_{x_1}(x_1,\theta) = -\alpha_2\delta + \beta_2(1-\delta) + \theta\gamma_{21}(\delta+\Delta) + 2\theta\gamma_{22}(\delta+\Delta)(X-\delta X+\delta x_1+\Delta x_1)$= (c). Both (a) and (c) are linear functions in θ. But when we couple these two segments of $C_{x_1}(x_1,\theta)$ together, we have $C_{x_1}(x_1,\theta)$ = min [(a), (c)]. The coupling through minimization relationship means that the $C_{x_1}(x_1,\theta)$ function is concave in θ. Thus, the K-irr case of Model 2 has a "learning effect" that pushes towards higher x_1 ($x_1^L \leq x_1^{NL}$).

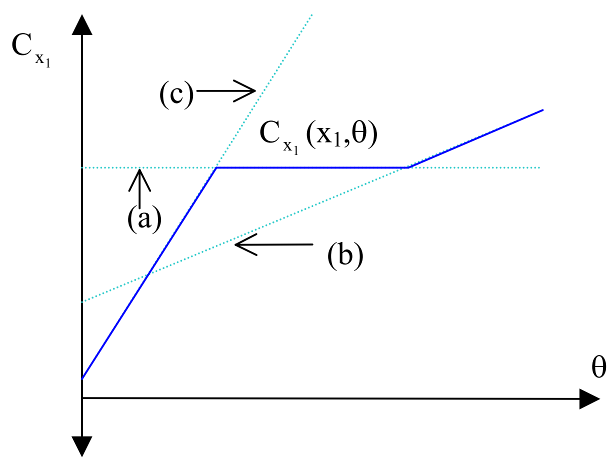


Figure 1d: G-irr, K-irr Case in Model 2

Figure 1d shows the graph of $C_{x_1}(x_1,\theta)$ for Model 2 when there are both G-irr and K-irr constraints. When $\theta < \dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}(X-\delta X+\delta x_1+\Delta x_1)}$ or when K-irr constraint is active, we have $C_{x_1}(x_1,\theta)$ = (c). When $\dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}(X-\delta X+\delta x_1+\Delta x_1)} \leq \theta \leq \dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}\Delta x_1}$ or when none of the constraints is active, the optimal solution is the interior solution and we have $C_{x_1}(x_1,\theta)$ = (a). When $\theta > \dfrac{(\alpha_2+\beta_2)}{\gamma_{21}+2\gamma_{22}\Delta x_1}$ or when G-irr is active, we have $C_{x_1}(x_1,\theta)$ = (b). Similar to the above cases, all the three functions are linear and the coupling of segment (a) and segment (b) occurs through maximization relationship while the coupling of segment (a) and segment (c) occurs through minimization relationship. Thus, the resultant $C_{x_1}(x_1,\theta)$ is neither concave nor convex in θ and we have the "learning effect" in the G-irr, K-irr case that is ambiguous in its direction. In other words, we may have $x_1^L > x_1^{NL}$ or $x_1^L = x_1^{NL}$ or $x_1^L < x_1^{NL}$ depending on the detailed parameters of the cost functions in the model as well as the probability distribution function of the uncertain parameter θ.

Table 3 summarizes the directions of the learning effects for all the four cases of Model 2.

Table3: Direction of the Learning Effect for Model 2

Irreversibility Assumption	**Curvature of C_{x1}**	**Direction of the "Learning Effect"**
1. No constraint	Linear in θ	None
2. With G-irr constraint	Convex in θ	Push towards lower x_1
3. With K-irr constraint	Concave in θ	Push towards higher x_1
4. With both G-irr and K-irr	Neither Convex Nor Concave	Ambiguous

<u>Discussion</u>

The analysis of the learning effect in Model 2 that we have done so far discusses each of the four cases separately. Alternatively, we may consider them as a group of related cases of the same model. Doing so provides an interesting perspective on the results.

First, note that Model 2 has no learning effect when there is no irreversibility constraint in the model. This is because when there is no irreversibility constraint or when none of the constraint(s) is active, $C_{x_1}(\cdot)$ becomes segment (a) which is linear in θ.

Second, when there is/are irreversible constraint(s) in the model, the function $C_{x_1}(\cdot)$ is composed of two major parts:

- A part where no constraint is active, i.e., segment (a).
- Parts where one of the irreversibility constraints is active, i.e., segment (b) where G-irr constraint is active and/or segment (c) where K-irr constraint is active.

Thus, the curvature of $C_{x_1}(\cdot)$ of any irreversible case is dependent on three elements: (i) the curvature of segment (a) where no constraint is active; (ii) the curvature(s) of the constrained segment(s), i.e., segments (b) and/or (c); and (iii) the couplings of various segments into the $C_{x_1}(\cdot)$ function.

Third, let us start by considering the curvatures of the constrained segments. According to the results above, both constrained segments – i.e., segments (b) and (c) – in Model 2 are found to be linear in θ. The linearity of these segments, however, is not limited to this model. Under the assumption of multiplicative uncertainty, the linearity of these two segments is due solely to the linearity of their corresponding irreversibility constraints on x_1. (Recall that the G-irr constraint requires that $x_2 \geq 0$ while K-irr constraint requires that $x_2 \leq X-\delta X+\delta x_1$.) Since all the models in this paper assume the same linear irreversibility constraints, the two segments are linear in all the other models as well.

This means that the curvature of $C_{x_1}(\cdot)$ in all the models in this paper is dependent on (i) the curvature of segment (a) and (ii) the coupling relationships of various segments in the $C_{x_1}(\cdot)$ function. As discussed below, this finding is very critical to the understanding of the concept of "learning effect" and "irreversibility bias on the learning effect".

Fourth, regarding to the curvature of the unconstrained segment (a). According to the results above, segment (a) of $C_{x_1}(\cdot)$ for Model 2 is also found to be linear. The linearity of the segment (a), however, is model specific. The curvature of segment (a) depends on the functional assumptions of each model.

Fifth, since all the three possible segments of $C_{x_1}(\cdot)$ in Model 2 are linear in θ. The curvature of $C_{x_1}(\cdot)$ in all irreversible cases of Model 2 is determined solely by the coupling relationships – the couplings of segment (a) with the other two segments of $C_{x_1}(\cdot)$. The coupling of segment (b) where the G-irr constraint is active with segment (a) where no constraint is active pushes $C_{x_1}(\cdot)$ function towards convexity (Figure 1b) – thus causing the G-irr case of Model 2 to have learning effect that pushes towards lower x_1. The coupling of segment (c) where the K-irr constraint is active with segment (a) pushes $C_{x_1}(\cdot)$ function towards concavity (Figure 1c) – thus causing the K-irr case of Model 2 to have learning effect that pushes towards higher x_1. Moreover, the coupling of both segments (b) and (c) with segment (a) in the G-irr, K-irr case causes the $C_{x_1}(\cdot)$ function to be neither concave nor convex (Figure 1d), thus we have the learning effect that has ambiguous direction.

The above results can be interpreted as if the addition of either of the two irreversibility constraints has created some kind of "irreversibility bias on the learning effect" that pushes the learning effect towards either direction. Since Model 2 has no learning effect when there is no constraint, the "learning effect" of each of the other three (irreversible) cases simply follows the direction(s) of the "irreversibility bias(es)" presented in the model. For the G-irr only case, since the coupling of segments (b) and (a) occurs through maximization relationship, the $C_{x_1}(\cdot)$ function is pushed towards convexity, and thus, we have an "irreversibility bias" that pushes the "learning effect" towards lower x_1 direction. Similarly, for the K-irr only case, since the coupling of segments (c) and (a) occurs through minimization relationship, the $C_{x_1}(\cdot)$ function is pushed towards concavity, and thus, we have an "irreversibility bias" that pushes the "learning effect" towards higher x_1 direction. Finally, when we have both G-irr and K-irr constraints in the model, we have two "irreversibility biases" that pushes against each other, thus we have the "learning effect" that has ambiguous direction.

This new perspective suggests that the directions of learning effects in all these cases are governed by the directions of three underlying components – i.e., the learning effect when no constraint, the irreversibility bias caused by G-irr constraint, and the irreversibility bias caused by K-irr constraint. Table 4 shows the directions of these three components for Model 2. A more precise definition of the irreversibility bias is given below.

Table 4: Directions of Underlying Components for Model 2

Learning Effect and Irreversibility Bias	**Direction**
Learning effect when no constraint	None
G-irreversibility bias	Push towards lower x_1
K-irreversibility bias	Push towards higher x_1

Definition of the "Irreversibility Bias on the Learning Effect"

The analysis of the learning effect of Model 2 above illustrates how the "learning effect when there is no constraint" combines with the "irreversibility bias(es) on the learning effect" in determining the direction of the "learning effect" in any irreversible case. It is important to note that:

- The direction of the "learning effect when there is no constraint" is governed by the concavity/convexity of the segment (a) where no constraint is active. (In the case of Model 2, segment (a) is linear in θ thus there is no "learning effect" when there is no constraint).
- The direction of the "irreversibility bias" is determined by the coupling relationship between segment (a) and the segment where the irreversibility constraint is active. Since both irreversibility constraints are linear in x_1 (G-irr: $x_2 \geq 0$, K-irr: $x_2 \leq X-\delta X+\delta x_1$), the segments of $C_{x_1}(\cdot)$ where these constraints are active become linear in θ. Thus the direction of the "irreversibility bias" is determined solely from the coupling relationship.

Definition 3 summarizes the definition of the "irreversibility bias on the learning effect".

For models with multiplicative uncertainty, the "irreversibility bias on the learning effect" of any linear irreversibility constraint can be defined as follows:

> *Definition 3.1* *The "irreversibility bias on the learning effect" pushes towards lower x_1 if and only if the coupling of the segment of $C_{x_1}(\cdot)$ where the irreversibility constraint is active with the segment of $C_{x_1}(\cdot)$ where no constraint is active occurs through maximization relationship.*
> *Definition 3.2* *The "irreversibility bias on the learning effect" pushes towards higher x_1 if and only if the coupling of the segment of $C_{x_1}(\cdot)$ where the irreversibility constraint is active with the segment of $C_{x_1}(\cdot)$ where no constraint is active occurs through minimization relationship.*

Analysis for Model 4 (I-quad, D- and O-linear)

Model 4 assumes that $I_t(\cdot)$is a quadratic function while $O_t(\cdot)$ and $D_t(\cdot)$ are linear functions. Under such assumptions, the objective functions when there is full learning and when there is no learning become:

$$
\begin{aligned}
J_L \quad &= \underset{x_1}{Min}\,[\alpha_1(X-x_1)+\beta_{11}(X-x_1)+\beta_{12}(X-x_1)^2+E\,\theta\gamma_1x_1+E\,[\underset{x_2}{Min}\,[\alpha_2(X-x_2)+\beta_{21}(x_1-x_2) \\
&\qquad +\beta_{22}(x_1-x_2)^2+\theta\gamma_2(x_2+\Delta x_1)]]] \\
&= \underset{x_1}{Min}\,[V^4(x_1, E\,\theta)+E\,C^4(x_1,\theta)].
\end{aligned}
$$

$$
\begin{aligned}
J_{NL} &= \underset{x_1}{Min}\,[\alpha_1(X-x_1)+ \beta_{11}(X-x_1)+ \beta_{12}(X-x_1)^2+ E\,\theta\gamma_1 x_1+ \underset{x_2}{Min}\,[\alpha_2(X-x_2)+ \beta_{21}(x_1-x_2)+ \beta_{22}(x_1-x_2)^2 \\
&\qquad + E\,\theta\gamma_2(x_2+\Delta x_1)]] \\
&= \underset{x_1}{Min}\,[V^4(x_1, E\,\theta) + C^4(x_1, E\,\theta)].
\end{aligned}
$$

Figure 2a and 2b show the graphs of $C_{x_1}(x_1,\theta)$ on the θ axis for the case of G-irr only and K-irr only, respectively. Only the graphs of $C_{x_1}(x_1,\theta)$ for these two cases are presented here since they carry sufficient information to allow the identification of the directions of the "learning effects" in all the four cases as well as the directions of all the three underlying components.

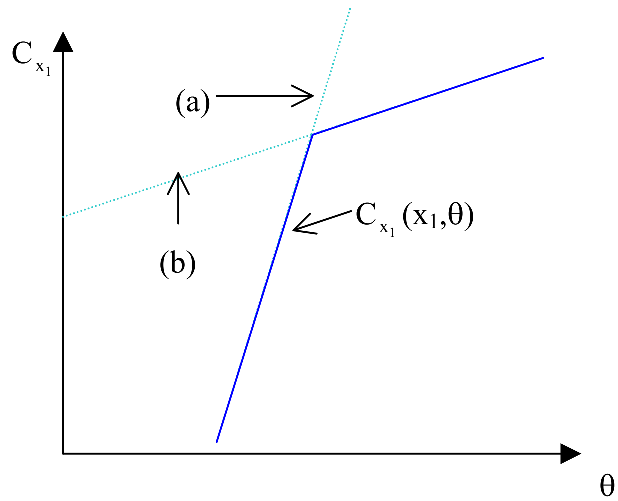


Figure 2a: G-irr Case in Model 4

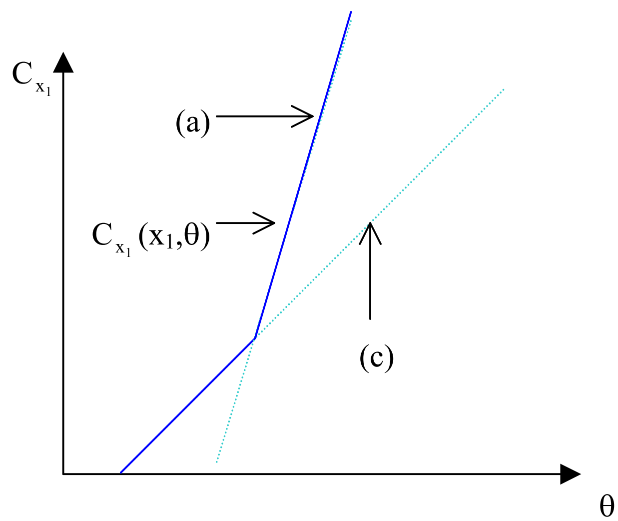


Figure 2b: K-irr Case in Model 4

Table 5 concludes the directions of the learning effects in all the four cases of Model 4. Table 6 summarizes the directions of the three underlying components that govern the directions of all the learning effects.

Table5: Direction of the Learning Effect for Model 4

Irreversibility Assumption	Curvature of C_{x1}	Direction of the "Learning Effect"
1. No constraint	Linear in θ	None
2. With G-irr constraint	Concave in θ	Push towards higher x_1
3. With K-irr constraint	Convex in θ	Push towards lower x_1
4. With both G-irr and K-irr	Neither Convex Nor Concave	Ambiguous

Table 6: Directions of Underlying Components for Model 4

Learning Effect and Irreversibility Bias	Direction
Learning effect when no constraint	None
G-irreversibility bias	Push towards higher x_1
K-irreversibility bias	Push towards lower x_1

Discussion

The analysis shows that, similar to the case of Model 2, there is no "learning effect when no constraint" in Model 4 – since the unconstrained segment (a) is linear in θ. As discussed earlier, the curvature/linearity of segment (a) found in these cases are due to the functional assumptions of the model. Actually, it can be shown that the unconstrained segment (a) of the C_{x_1} function is always linear in θ for all linear and single non-linear models.

For the G-irr bias, the coupling of segment (b) where the G-irr constraint is active with segment (a) occurs through minimization relationship (Figure 2a). This causes the $C_{x_1}(\cdot)$ function to be pushed towards convexity. Therefore, the incorporation of G-irr constraint in the model creates the G-irr bias that pushes the learning effect towards higher x_1 direction. On the other hand, the coupling of segment (c) where the K-irr constraint is active with segment (a) occurs through maximization relationship (Figure 2b). The $C_{x_1}(\cdot)$ function is pushed towards convexity, and thus, we have the K-irr bias that pushes the learning effect towards lower x_1 direction. So, similar to the case of Model 2, these two irreversibility constraints create irreversibility biases that push against each other. But, contrary to Model 2, the directions of the two biases are flipped. In Model 2 we have G-irr bias pushing towards lower x_1 and K-irr bias pushing towards higher x_1, while in Model 4 we have G-irr bias pushing towards higher x_1 and K-irr bias pushing towards lower x_1.

The directions of the irreversibility biases found in Model 4 seem counter-intuitive. It looks like the two irreversibility biases found in this model are pushing the optimal solution towards less flexibility choices – G-irr bias pushing towards more emissions while K-irr bias pushing towards less emissions. These irregularities can be explained in part by the insight gain from understanding the "Simple Analysis 1" approach discussed below.

8. Simple Analysis 1

Drawing from the results of the analysis, I find that for linear and single non-linear models there is a much simpler and more intuitive way to reach the same results. This intuitive approach – called Simple Analysis 1 (SA1) – is possible due to a special characteristic of the linear and single non-linear models. By selecting the control variable and rearranging the model appropriately, all linear and single non-linear models can be rewritten such that they are functionally independent. This characteristic allows us to intuitively analyze and understand the “irreversibility effect” in these models by simply looking at the constraint dependency mechanism of the rearranged models. The most important step of SA1, thus, is the selection of the appropriate choice of control variable(s).

Consider Model 2 first. Using G_t as the control variable, the objective function of Model 2 can be written as follows:

$$
\begin{aligned}
J \quad &= \underset{G_1}{Min}\ [\alpha_1(X-G_1)+\beta_1(X-G_1)+\theta\gamma_{11}G_1+\theta\gamma_{12}G_1^2+\underset{G_2}{Min}\ (\alpha_2(X-G_2+\Delta G_1) \\
&\quad +\beta_2(G_1-G_2+\Delta G_1)+\theta\gamma_{21}G_2+\theta\gamma_{22}G_2^2)] \\
&= \underset{G_1}{Min}\ [\alpha_1(X-G_1)+\alpha_2\Delta G_1+\beta_1(X-G_1)+\beta_2(G_1+\Delta G_1)+\theta\gamma_{11}G_1+\theta\gamma_{12}G_1^2 \\
&\quad +\underset{G_2}{Min}\ (\alpha_2(X-G_2)-\beta_2G_2+\theta\gamma_{21}G_2+\theta\gamma_{22}G_2^2)].
\end{aligned}
$$

Since the objective function can be rearranged such that it is functionally independent, G_t is an appropriate control variable choice for Model 2. Please note that if we rewrite Model 2 using some other choices of control variable (e.g., x_t or Q_t), the model may not be functionally independent.

Epstein (1980) has proved – subject to standard qualifications – that when a model is functional independent (i.e., additively separable), the "irreversibility effect" will always hold. Since Model 2 can be written such that it is functionally independent based on variable G_t, the "irreversibility effect" musts hold for Model 2 providing that we base our definition of "irreversibility effect" on how the constraints can be interpreted in terms of variable G_t.

What this conclusion actually means, in terms of our new framework, is that: (i) There is no "learning effect when no constraint" in Model 2 and, thus, the "learning effects in irreversible cases" of this model are solely the results of the "irreversibility biases" presented in the model. And (ii) the directions of the "irreversibility biases" in Model 2 are such that they always push the optimal current-period solution towards more flexible choices – providing that we base our definition of more/less flexibility on how the constraints can be interpreted in terms of the appropriate variable G_t.. The same is true for other linear and single non-linear models. Thus, for this group of models, we can determine the direction of the "irreversibility bias" of any constraint by looking at how the constraint can be interpreted in terms of the appropriate control variable(s) of that model.

In terms of variable G_t, the G-irr constraint requires that we choose $G_2 \geq \Delta G_1$. This constraint imposes a lower limit on the permissible choices of G_2 and this lower limit is set to be an increasing function of G_1. This means that when the value of G_1 is higher, the lower constraint will become higher and, thus, less flexible. Therefore the G-irr constraint is found to create an "irreversibility bias" that would bias our choice towards lower G_1 – i.e., lower current period emissions (lower x_1).

For the K-irr constraint, in terms of variable G_t, the K-irr constraint requires that $G_2 \leq Z-\delta Z+\delta G_1+\Delta G_1$. This constraint imposes an upper limit that is increasing in G_1 on the permissible choice of G_2. This means that the higher the value of G_1, the larger the upper limit will be. Since a larger upper limit means a more flexible limit, the K-irr constraint would create an "irreversibility effect" that pushes towards higher current period emissions (higher G_1 and x_1).

The directions of the two irreversibility biases are consistent with the conventional conception held in most "irreversibility effect" literature. Since the G-irr constraint limits us from unemitting the past emissions, it should push us towards emitting less now. Because we do not want to be stuck with a high stock of atmospheric GHGs if the damage of climate change turns out to be severe. Analogously, for the case of K-irr constraint, since we do not want to be stuck with high stock of GHG abatement capital that could not be uninvested if the climate change damages turns out to be minimal, the existence of K-irr constraint push us towards investing less (or emitting more) now.

Next, consider Model 4. Model 4 can be rearranged such that it becomes functionally independent when a new variable Q_t is selected as the control variable. The variable Q_t is defined as the difference in abatement level compared to the previous period ($Q_t = A_t - A_{t-1}$) – i.e., it reflects the level of abatement capital investment during period t. Using Q_t as the control variable, the objective function of Model 4 can be rearranged such that it becomes functionally independent.

$$
\begin{aligned}
J &= \underset{Q_1}{Min}[\alpha_1 Q_1 + \beta_{11}Q_1 + \beta_{12}Q_1^2 + \theta\gamma_1(X-Q_1) + \underset{Q_2}{Min}[\alpha_2(Q_1+Q_2) + \beta_{21}Q_2 + \beta_{22}Q_2^2 \\
&\qquad + \theta\gamma_2((1+\Delta)(X-Q_1)-Q_2)]] \\
&= \underset{Q_1}{Min}[\alpha_1 Q_1 + \alpha_2 Q_1 + \beta_{11}Q_1 + \beta_{12}Q_1^2 + \theta\gamma_1(X-Q_1) + \theta\gamma_2(1+\Delta)(X-Q_1) + \underset{Q_2}{Min}[\alpha_2 Q_2 \\
&\qquad + \beta_{21}Q_2 + \beta_{22}Q_2^2 - \theta\gamma_2 Q_2]]
\end{aligned}
$$

For the G-irr constraint, in terms of variable Q_t, the G-irr assumption requires that $Q_2 \leq X-Q_1$. It imposes an upper limit constraint on the choice of Q_2 and this upper limit is set to be a decreasing function of Q_1. This means that when the value of Q_1 gets higher, the upper limit becomes smaller and less flexible. Thus, for Model 4, the G-irr constraint would create an irreversibility bias that pushes our current choice toward lower Q_t (lower investment), i.e., higher x_1 (higher emissions).

For the K-irr constraint, the constraint requires that $Q_2 \geq -(1-\delta)Q_1$. It imposes a lower limit constraint on the choice of Q_2 and this constraint is decreasing in Q_1. Thus, under the K-irr assumption, an increase in Q_1 will make the lower constraint smaller, i.e., more relaxed. Therefore, the K-irr assumption would create an irreversibility bias that pushes towards higher Q_1. But this means that the K-irreversibility bias is pushing towards lower current emissions (lower x_1).

The results of Model 4 may seem counter-intuitive, still they can be readily understood. Starting with the G-irr constraint, the constraint says that we cannot emit a negative amount of emissions ($x_2 \geq 0$) or that we cannot abate more than the amount that we would emit when there is no abatement ($A_2 \leq X$). The upper limit on the amount of abatement we can do means that there is an upper limit on the amount of investment we can add in each period. Surprisingly, the more abatement we have in the current period, the less room we leave for future decision on the additional investment. Thus when the G-irr assumption is translated into the restriction in Q_2 ($Q_2 \leq X\text{-}Q_1$) the constraint would become stricter when there is higher abatement investment in period one (higher Q_1). Therefore, we find the G-irr constraint creating an "irreversibility effect" that push towards lower Q_1 (higher x_1) in the model that has Q_t as the appropriate choice of control variable.

Whereas the keyword in understanding the results in G-irr case is “investment”, the keyword for the K-irr case is "capital depreciation". Under the assumption of K-irr, even though it is not possible to uninvest the abatement capital, there is still a small possibility that we can have $A_2 > A_1$ ($Q_2 < 0$) – by allowing the existing capital to depreciate. The K-irr constraint specifies the minimum amount of GHG abatement such that all the existing capital after depreciation must be utilized ($A_2 \geq \delta A_1$). This is equivalent to, in terms of capital depreciation, specifying the maximum capital depreciation that could happen. Thus, the more investment we have made in the past, the more options we have in deciding how much existing abatement capital we would allow to decay. Therefore when the K-irr assumption is considered in terms of the restriction in Q_t, it would create an "irreversibility effect" that biases us toward more current investment (less emissions).

9. Conclusion

This paper proposes a new framework in understanding and analyzing an intertemporal interaction between learning and decision making processes under the context of uncertainty and irreversibility. It proposes that this intertemporal feature, conventionally termed as the “irreversibility effect”, should be renamed as the “learning effect” – because the focus of the analysis of the “learning / irreversibility effect” is the effect of future learning, not the effect of irreversibility, on the optimal current-period solution. The paper also introduces another new definition of “irreversibility bias on the learning effect” that reflects how the addition of the irreversibility constraint into the model affects the direction of the “learning effect”.

The analysis has been conducted on four simple models of climate change problem that differ in their functional assumptions. The analysis illustrates how the “learning effect when no constraint” and the “irreversibility bias(es)” work in creating the learning effect in the irreversible case under the new framework. The results of the analysis as well as the use of SA1 approach point to the critical significance of the functional assumptions in determining the “learning effect” and the “irreversibility bias” in the model.

References

Arrow, K. and A. C. Fisher (1974). "Environmental preservation, uncertainty, and irreversibility." Quarterly Journal of Economics **88**(2): 312-319.

Epstein, L. G. (1980). "Decision-Making and the Temporal Resolution of Uncertainty." International Economic Review **21**(2): 269-283.

Freixas, X. and J.-J. Laffont (1984). On the irreversibility effect. Baysian Models in Economic Theory. M. Boyer and R. E. Kihlstrom, Elsevier Science Publisher B.V.: 105-114.

Gollier, C., B. Jullien, and N. Treich (2000). "Scientific progress and irreversibility: an economic interpretation of the 'Precautionary Principle'." Journal of Public Economics **75**(2): 229-253.

Henry, C. (1974). "Investment decisions under uncertainty: The irreversibility effect." American Economic Review **64**: 1006-1012.

Kolstad, C. D. (1996). "Fundamental irreversibilities in stock externalities." Journal of Public Economics **60**(2): 221-233.

Narain, U., M. Hanemann, and A. C. Fisher (2003). Uncertainty, learning, and the irreversibility effect.

Ulph, A. and D. Ulph (1995). Global warming, irreversibility and learning. Discussion Papers in Economics and Econometrics 9601. University of Southampton.

Ulph, A. and D. Ulph (1997). "Global warming, irreversibility and learning." Economic Journal **107** (442): 636-650.