Title
The irreversibility effect in environmental decisionmaking

Journal
Environmental & Resource Economics, 38(3)

ISSN
1573-1502

Authors
Narain, Urvashi
Hanemann, W. Michael
Fisher, Anthony C

Publication Date
2007-11-01

DOI
10.1007/s10640-007-9083-x

Peer reviewed

The Irreversibility Effect in Environmental Decisionmaking

Urvashi Narain
Resources for the Future
1616 P Street NW, Washington DC 20036

Michael Hanemann
Department of Agricultural & Resource Economics
University of California, Berkeley CA 94720

Anthony Fisher
Department of Agricultural & Resource Economics
University of California, Berkeley CA 94720

This research received financial support from the National Science Foundation grant No. SES-9818642. The authors gratefully acknowledge comments on earlier drafts of this paper by an anonymous reviewer, Christian Gollier, Christian Traeger, an anonymous reviewer, and participants at the International Conference on Risk and Uncertainty in Environmental and Resource Economics, Wageningen, The Netherlands, and Columbia Earth Institute Environmental Economics Seminar.
Corresponding author: Anthony Fisher, Department of Agricultural and Resource Economics, University of California, Berkeley, CA 94720, Tel: (510) 642-7555, Fax: (510) 643-8911, Email: fisher@are.berkeley.edu.

The Irreversibility Effect in Environmental Decisionmaking

Abstract

We provide a new, more general, definition for the irreversibility effect and demonstrate its relevance to problems involving environmental and other decisions under uncertainty. We establish several analytical and numerical results that suggest both that the effect holds more widely than generally recognized, and that an existing result (Epstein's Theorem), giving a sufficient condition for determining whether the effect holds, can be applied more widely than previously indicated, in particular to problems involving intertemporally nonseparable benefit functions. We further show that a low elasticity of intertemporal substitution will however result in failure of the effect.

JEL: Q20, Q30, Q51

Keywords: decisionmaking under uncertainty, irreversibility effect, necessary and sufficient conditions, nonseparable benefit functions

1. Introduction

Environmental impacts of an investment in resource development can be long lasting, or even irreversible. This is a feature of environmental valuation and decision problems that has received a great deal of attention in the literature, based on findings in the natural sciences. For example, there is both scientific and popular concern today about loss of biodiversity, the genetic information that is potentially valuable in medicine, agriculture, and other productive activities. Much of the concern is for endangered species, or their habitats such as tropical moist forests that are subject to more or less irreversible conversion to other uses. But even if species survival is not at issue, biological impacts can be very difficult to reverse over any relevant time span. The clear-cutting of a climax forest species, for example, removes the results of an ecological succession that may represent centuries of natural processes. Regeneration may not lead to the original configuration, as opportunistic species such as hardy grasses come in and preempt the niche otherwise filled by the climax species (Albers & Goldbach 2000).

Irreversibilities have also been identified as a key feature of the problem of how to respond to potential impacts of climate change. Emissions of greenhouse gases, in particular carbon dioxide, accumulate in the atmosphere and decay only slowly. According to one calculation, assuming business-as-usual use of fossil fuels over the next several decades, after a thousand years carbon dioxide concentrations will still be well over twice the current level, and nearly three times the pre-industrial level, and will remain elevated for many thousands of years (Schultz & Kasting 1997). There is also some prospect of essentially irreversible catastrophic impact as would result for example from the disintegration of the West Antarctic Ice Sheet and consequent rise in sea level of 15-20 feet. Recent findings suggest that this possibility is more serious, and perhaps closer in time, than economists (and others) have realized ((Kerr 1998) and (de Angelis & Skvarca 2003)).

Irreversibilities are of course not confined to environmental decisions, but occur in a wide variety of economic settings, as the definitive work on investment decisions under uncertainty by Dixit & Pindyck (1994) makes clear.

In the environmental economics literature the analysis of investment decisions under uncertainty and irreversibility was introduced by Arrow & Fisher (1974) and Henry (1974), who show that, for a linear net benefit function or an all-or-nothing choice, it will be optimal to delay or reduce investment, for example in a water resource development project in a natural environment, if future net benefits are uncertain, investment decisions are irreversible, and there is a possibility of learning about future benefits. Dixit and Pindyck and others establish essentially the same result for the more general investment problem, broadening the treatment to include nonlinear benefit functions and continuous choices, at the same time greatly enriching the analysis with a rigorous treatment of stochastic optimization.

Beginning with the seminal paper by Epstein (1980) on decision-making and the temporal resolution of uncertainty, and including important contributions by Freixas & Laffont (1984), Hanemann (1989), Kolstad (1996), Ulph & Ulph (1997), Gollier, Jullien & Treich (2000), and Ha-Duong & Treich (2004), another strand of the literature has focused on the question of whether the rather strong and unambiguous results of Arrow and Fisher, Henry, and Dixit and Pindyck, continue to hold in still more general settings in which the intertemporal benefit function exhibits properties not considered by these authors.

In this paper we take up the discussion of several aspects of this question. The next section reviews existing definitions of the irreversibility effect and proposes a new one, which we show in Section 3 to be more generally applicable. Section 4 develops a numerical example to prove that the only necessary condition in the literature is, in fact, sufficient but not necessary, while Section 5 establishes that one of the mostly widely used sufficient conditions in the literature, due to Epstein (1980), is more widely applicable than previously believed. This section also establishes conditions under which the irreversibility effect is likely to be violated. Section 6 offers some broad conclusions on the status and significance of the irreversibility effect.

2. The Irreversibility Effect and the Environment

We begin by presenting two existing definitions, and one new definition, for the irreversibility effect. For this purpose consider a two-period decision problem, where in the first period the decision maker chooses a variable x_1 and in the second period a variable x_2. Net benefits in the first period, denoted by $B_1(x_1)$, are deterministic and depend only on x_1, but net benefits in the second period, denoted by $B_2(x_1, x_2, z_i)$, are stochastic and are a function both of x_1, x_2, and also of z, a random variable that reflects the underlying *uncertainty* about the nature of net benefits. We assume that z is a discrete random variable with M possible realizations. Furthermore, B_1 is assumed to be concave and twice continuously differentiable in x_1, and B_2 concave and twice continuously differentiable in x_1 and x_2. An issue that will become of some importance is whether or not the benefit function is separable in x_1 and x_2. In the general case where B_2 is a function of x_1, the benefit function is said to be nonseparable. If, on the other hand, B_2 were only a function of x_2 and z but not of x_1, then the benefit function would be said to be separable.

In principle, there are constraints on the first- and second-period choices. C_1 denotes the constraint function for x_1. A crucial issue in the literature is the extent to which the first period choice of x_1 constrains the future choice of x_2. In general, we will assume that the first period choice does constrain the second period choice, the constraint on the latter being given by $C_2(x_1)$. The constraint on x_2 could take a variety of forms and, in general, it implies a loss of flexibility in the second period decision. A sharp form of the constraint would be $x_1 > x_2$ which implies that x_2 is constrained to be less than x_1; we refer to this, and any such constraint on x_2, as the *irreversibility* constraint. Note that, by using a non-separable formulation of the second period net benefit function, we already imply that the first period decision will affect the choice confronting the decision maker in the second period. Making the second period constraint function depend on x_1 introduces a separate element of interdependence between the two choices.

Before the second period decision is made, the decisionmaker receives a signal, denoted by y_j, that reveals some information about z. This is the source of *learning.* y is also assumed to be a discrete random variable with N possible realizations. The amount of information contained in y depends on how closely related z and y are. Let y and y' denote two potential signals where the correlation between y and z is greater than the correlation between y' and z. y is said to be more informative about z, and leads to greater learning about the true nature of z, than y'.[1] After the signal is received, the decisionmaker updates her prior expectations about z by formulating a posterior distribution denoted by $\pi_{ij} = p(z = z_i/y = y_j)$ and then chooses x_2 for each signal to maximize the expected benefit over the different states. Also, let q_j denote the probability distribution for y.

With this notation, the dynamic optimization problem is

$$(1) \qquad \max_{x_1 \in C_1} \left(B_1(x_1) + \sum_j q_j \max_{x_2 \in C_2(x_1)} \left[\sum_i \pi_{ij} B_2(x_1, x_2, z_i) \right] \right).$$

Finally, we assume that a unique solution exists, and lies in the interior of C_1. Let x_1^* denote the maximum corresponding to the more informative signal y, and x_1^{**} the maximum corresponding to the less informative signal y'.

The conventional definition of the irreversibility effect in the literature is

$$(2) \qquad \text{either } x_1^* \geq x_1^{**} \textit{ or } x_1^* \leq x_1^{**}.$$

Which of these two conditions applies depends on the structure of the problem. If the problem is how much wildlife habitat to keep intact in the first period and not convert to farmland, then an increase in the fraction of habitat left untouched in anticipation that the decisionmaker will learn about the relative benefits of wildlife habitats and farmlands prior

[1]Note that we are adopting the same notion of learning as adopted by Epstein (1980), who, in turn, uses the notion of greater information discussed by Marschak & Miyasawa (1968). For a more precise definition of greater learning see Epstein (1980).

to the second period, as compared to the amount of habitat left untouched when there is no possibility of learning, implies an irreversibility effect. Converting a larger fraction of habitat into farmland in the first period, before uncertainty about the benefits of keeping wildlife habitats intact is resolved, would force the decisionmaker to accept lower benefits should it turn out that the benefits of habitats are larger than initially expected. If the benefits are smaller than expected then the decision maker can choose to convert more habitat into farmland in the second period, a possibility that is in no way constrained by leaving a larger fraction of the land as wildlife habitat in the first period. In this case the irreversibility effect holds if $x_1^* \geq x_1^{**}$, where x_1 is the amount of land in wildlife habitat in the first period.

On the other hand, if the decision is how much of a greenhouse gas to emit when damages due to global warming are uncertain, then a decrease in the amount emitted implies an irreversibility effect. Higher first period emissions would lock the decisionmaker into accepting whatever the nature of the damages are revealed to be, and not being able to avert damages should these turn out to be higher than expected. Should damages turn out to be lower than expected, the decisionmaker can always increase emissions in the second period. In this case the irreversibility effect holds if $x_1^* \leq x_1^{**}$, where x_1 is the amount of the greenhouse gas emitted in the first period.

A related definition, since the conventional definition is really about allowing for more options in the future, due to Freixas & Laffont (1984), is

$$C_2(x_1^*) \supseteq C_2(x_1^{**}). \tag{3}$$

In words, the irreversibility effect is said to hold if the second period choice set associated with x_1^* is at least as large as the choice set associated with x_1^{**}. In the habitat versus farmlands example, where the decisionmaker chooses the amount of wildlife habitat to leave intact in the second period and where $C_2(x_1)$ is defined as $x_1 \geq x_2$, the second period choice set is larger the larger is the amount of habitat left intact in the first period. The irreversibility effect is said to hold if $x_1^* \geq x_1^{**}$, which in turn is equivalent to $C_2(x_1^*) \supseteq C_2(x_1^{**})$.

We propose a third, more general definition. Define $\hat{x}_1$ as the value of x_1 that gives maximum decisionmaking flexibility in the future. For example, if x_2 is constrained to be greater than (less than) x_1, $x_1 \in [0,1]$ and $x_2 \in [0,1]$, then $\hat{x}_1 = 0$ ($\hat{x}_1 = 1$). This is because with $x_1 = 0$ ($x_1 = 1$) there is no constraint on the choice of x_2, and so there is maximum decisionmaking flexibility. In terms of the second period choice set $\hat{x}_1 = 0$ ($\hat{x}_1 = 1$) implies a set that consists of all possible values of the second period choice variable, x_2. We will say that an irreversibility effect exists if

$$(4) \qquad |x_1^* - \hat{x}_1| \leq |x_1^{**} - \hat{x}_1|,$$

that is, if the optimum corresponding to the more informative signal is at least as close to the point of maximum flexibility as the optimum corresponding to the less informative signal. In some models $\hat{x}_1$ may be a constant, while in others it may be determined by the model parameters. The virtue of this definition is that it encompasses both cases under equation (2) and is therefore independent of the structure of the problem. It is equivalent to $x_1^* \geq x_1^{**}$ in cases where x_2 is constrained to be less than x_1, $x_1 \in [0,1]$ and $x_2 \in [0,1]$. In such cases, $\hat{x}_1 = 1$, and according to our definition, the irreversibility effect holds if $|x_1^* - 1| \leq |x_1^{**} - 1|$. Since x_1 lies between 0 and 1, this simplifies to $x_1^* \geq x_1^{**}$. Alternatively, if x_2 is constrained to be greater than x_1, $x_1 \in [0,1]$ and $x_2 \in [0,1]$, then since $\hat{x}_1 = 0$, our definition simplifies to $x_1^* \leq x_1^{**}$.

Another advantage of our definition is that it applies to problems where the two received definitions fail, namely when the decisionmaker chooses two *different* objects in the first and the second periods. In the farmland versus wildlife habitat example the decisionmaker chooses the same object in each period, namely the amount of habitat to leave intact. Similarly, in the global warming example the object chosen in each period is the amount of greenhouse gases to emit. However, in the case of Epstein's firm's-demand-for-capital model (discussed in detail in the next section) a different object is chosen in each period: capital in the first and labor in the second. Alternatively, faced with a decision to reserve a natural environment for wildlife habitat or recreation, a manager may choose the amount of land to set aside in the first period and the number of rangers and complementary personnel to allocate to monitoring and protection in the second period. In such models with two different choice variables, the choice made in the first period does not limit the choice set available in the second period. The amount of capital chosen in the first period does not limit the amount of labor that can be employed in the second period. The amount of land set aside for protection does not limit the number of personnel that can be employed. Consequently, a definition of irreversibility that compares the choice sets in the second period for different initial choices is not useful and will not yield any predictions about the irreversibility effect. The other received definition of the irreversibility effect, specified in equation 2, fails as well because it is not clear which of the two conditions specified under this definition applies to such models. Our definition, on the other hand, and as we show in the next section, continues to apply.

3. General Definition

To establish that our definition is more general we rely on a condition established by Epstein (1980) for the irreversibility effect to hold, a sufficient condition under which the initial level of investment in a two-period model with uncertainty and the possibility of future learning is less than the initial level with uncertainty and no or less learning. Using the model in Section 2, let $J(x_1, \xi_j)$ denote the value function associated with the signal y_j, and be defined as

$$J(x_1, \xi_j) \equiv \max_{x_2 \in C_2(x_1)} \sum_i \pi_{ij} B_2(x_1, x_2, z_i) \tag{5}$$

where $\xi_j = [\pi_{1j}, \pi_{2j}, ..\pi_{ij}.., \pi_{Mj}]$ and is a vector of the posterior probability distribution corresponding to the signal y_j. $J(x_1, \xi)$ is then a vector of value functions where $\xi = [\xi_1, \xi_2, ...\xi_j, ...\xi_N]$. Under the assumption that $J(x_1, \xi)$ is concave and differentiable with respect to x_1,[2] then Epstein's sufficient condition relating x_1^* to x_1^{**} is given in Theorem 1.

Theorem 1. *If $J_{x_1}(x_1^*, \xi)$ is a concave (convex) function of ξ, then $x_1^* \leq (\geq) x_1^{**}$. If $J_{x_1}(x_1^*, \xi)$ is neither convex nor concave, then the sign of $x_1^* - x_1^{**}$ is ambiguous.*

In words, the sufficient condition states that if the slope of the value function with respect to x_1, $J_{x_1}(x_1, \xi)$, is concave (convex) in the posterior probability distribution, then the optimal choice of x_1 associated with the more informative signal is less (more) than the optimal choice associated with the less informative signal.[3]

Now consider Epstein's firm's-demand-for-capital example. In this example, a firm chooses its investment in capital in the first period, and determines its demand for labor in the second, in order to maximize profits. Capital is thus quasi fixed while labor is variable and moreover, in the second period the firm can not disinvest its capital. Price of the output is unknown in the first period, but the firm receives some information about output prices at the beginning of the second period. The firm therefore solves the following problem:

$$\max_{K \geq 0} \left(-cK + \sum_j q_j \max_{L \geq 0} \Big(\sum_i \pi_{ij} p_i F(K, L) - wL \Big) \right) \tag{6}$$

where K denotes capital, L denotes labor, c is the cost of capital, w is the wage rate, F is a strictly concave production function and p_i is the unknown output price.

Since the first period choice, capital, enters the benefit function in the second period the benefit function is said to be intertemporally nonseparable. Note that if the firm was allowed

[2]This assumption holds if $B_2(x_1, x_2, z)$ is concave in x_1 and x_2 and if for $C_2(x_1) = \{x_2 | f(x_1, x_2) \geq 0\}$, the function f is concave (Epstein 1980).

[3]This sufficient condition is hard to relate to the primitives of an economic model. As it stands it is not clear what type of model gives rise to a concave or convex slope of the value function and thus qualifies for application of Epstein's Theorem. Gollier et al. (2000) provide necessary and sufficient conditions for two classes of models under which the second derivative of the slope of the value function can in fact be signed.

to invest or disinvest in capital in the second period then the problem faced by the firm would become intertemporally separable. Say, for example, that the firm is allowed to invest in the capital stock in the second period, though at a higher cost. The problem described by equation (6) would change to

$$\max_{K_1 \geq 0}\left(-c_1 K_1 + \sum_j q_j \max_{L\geq 0, K_2 \geq K_1} \big(\sum_i \pi_{ij} p_i F(K_2, L) - wL - c_2(K_2 - K_1)\big)\right) \tag{7}$$

where K_1 denotes capital in the first period, K_2 capital in the second period, c_1 is the cost of capital in the first period and c_2 the cost in the second period. Since capital is more costly in the second period, $c_2 > c_1$. Equation (7) can then be re-written as

$$\max_{K_1 \geq 0}\left((c_2 - c_1) K_1 + \sum_j q_j \max_{L\geq 0, K_2 \leq K_1} \big(\sum_i \pi_{ij} p_i F(K_2, L) - wL - c_2 K_2\big)\right).$$

Since the benefit function in the second period is no longer a function of K_1, the problem is intertemporally separable. A similar case can be made for when the firm is allowed to disinvest in the second period.

Coming back now to the nonseparable version of this problem, and according to Epstein's sufficient condition, whether the irreversibility effect holds in this problem depends on the second derivative of the slope of the value function in the random variable. For the following constant elasticity of substitution production function

$$F(K, L) = [aK^{-\beta} + bL^{-\beta}]^{\frac{-\mu}{\beta}}$$

where $a > 0$, $b > 0$, $\beta > -1$, $\beta \neq 0$, $0 < \mu < 1$ (μ being a measure of returns to scale) and the elasticity of substitution, σ, is equal to $\frac{1}{(1+\beta)}$, Hartman (1976) has established that the third derivative of the value function depends on the relationship between the elasticity of substitution and the returns to scale. Specifically, Hartman has shown that if $\sigma > (<) \frac{1}{(1-\mu)}$ then $J_K(K, p_i)$ is concave (convex) in p_i. This combined with Theorem 1 implies that if $\sigma > (<) \frac{1}{(1-\mu)}$ then the demand for capital is lower (higher) when there is a possibility of learning than when there is no possibility of learning. By implicitly relying on the existing definitions of the irreversibility effect, which, in turn imply that the effect holds either when $K^* \leq K^{**}$ or when $K^* \geq K^{**}$ (see equation 2), and since the demand for capital does not unambiguously increase or decrease with learning, Epstein leads the reader to conclude that the irreversibility effect is violated in this example. However, our definition of the irreversibility effect establishes that the effect does, in fact, hold in this example.

Observe that the firm can neither increase nor decrease its capital stock in the second period. Consequently, one cannot tell *a priori* whether a high or a low demand for capital in the first period constitutes a flexibility-enhancing decision, and therefore whether $K^* \leq K^{**}$ or $K^* \geq K^{**}$ is required for the irreversibility effect to hold. When σ is high so that capital and labor can be easily substituted then a *lower* capital stock today may very well give the decision maker greater flexibility tomorrow. If it turns out that the decision maker has underestimated his or her production needs, then he or she can compensate for the low stock of capital by hiring more labor. The irreversibility effect would then hold if $K^* \leq K^{**}$. On the other hand, if σ is low so that capital and labor cannot be substituted, a *higher* capital stock today may maintain greater flexibility tomorrow and the irreversibility effect would hold if $K^* \geq K^{**}$. Therefore in order to establish whether or not the irreversibility effect holds, we need to first define what constitutes flexibility in this problem.

There are two definitions of flexibility in the literature: one due to Freixas & Laffont (1984), and another due to Jones & Ostroy (1984). By Freixas and Laffont's definition of flexibility, which is equivalent to their definition for the irreversibility effect (see equation 3), the choice of capital that gives the greatest flexibility in the second period is the one that produces the largest choice set in the second period. Jones and Ostroy similarly define flexibility in terms of second period choices that can be attained from the first period position with the additional qualifier that the positions be attained at a given cost and for a particular state of the world.[4] Neither of these definitions apply to our problem, however. If the choice

[4]Let $c(x_1, x_2, z_i)$ denote the cost of moving from x_1 to x_2 given that the state of the world is z_i. Then $G(x_1, z_i, \alpha)$, where

$$G(x_1, z_i, \alpha) \equiv \{x_2 : c(x_1, x_2, z_i) \leq \alpha\},$$

set in the second period is defined in terms of capital, then since capital can neither be increased nor decreased in the second period, irrespective of the level of capital chosen in the first period, the decision maker has a single element in their choice set in the second period, namely the level of capital in the first period. No one level of capital gives a larger or smaller choice set in the second period. Defining the second period choice set in terms of labor instead does not help define a more or less flexible level of capital either as the first period's choice of capital in no way restricts the choice of labor in the second period.

The question then arises, with respect to what variable should flexibility be measured? So far we have tried to measure flexibility in terms of the choice variables, that is, in terms of the choices of capital or labor in the second period that are feasible given the choice of capital in the first period. However, one could instead measure flexibility in terms of the level of output that can be attained in the second period given the choice of capital in the first. After all, the firm cares about the level of capital, or any other input, only in so far as it allows the firm to produce output in the second period. In fact, what the firm really cares about is the *range of outputs*[5] that can be attained for a given level of capital. If the firm learns that the price of output is likely to be high tomorrow it would want to produce more, and conversely if it learns that the price is likely to be low it would want to reduce production. Flexibility for the firm manifests itself in terms of the range of output that the firm can produce. With this definition a more flexible level of capital is one that enables the firm to produce a greater range of output in the second period.

With a production function that exhibits constant elasticity of substitution, the level of capital that enables the firm to produce the greatest range of output, in fact, depends on the model parameters. Furthermore, the level of capital that gives the greatest flexibility is lower (higher) when $\sigma > (<) \frac{1}{1-\mu}$.

is the set of second period positions attainable from x_1 at a cost that does not exceed α in state s. In general x_1^* is said to generate more flexibility than x_1^{**} when for all $\alpha \geq 0$ and for all z_i, $G(x_1^*, z_i, \alpha) \supseteq G(x_1^{**}, z_i, \alpha)$.

[5]Note that this is consistent with Hirshleifer & Riley (1992) who point out that flexibility is different from the range of actions which in our example would mean the range of capital or labor. We instead equate flexibility to the range of outputs.

Proposition 1. *If* $\sigma > (<)\frac{1}{1-\mu}$ *then* $\hat{K} = \underline{K}(\overline{K})$.

where $\hat{K}$ is the level of capital that implies the greatest amount of flexibility, $\underline{K}$ is the minimum capital stock and $\overline{K}$ is the maximum capital stock.[6]

Proof. Let $\bar{y}(K)$ denote the range of output that can be achieved for a given level of capital and let $\gamma = \frac{-\mu}{\beta}$. Note that when $\sigma > (<)\frac{1}{1-\mu}$, $\gamma < (>)1$ since $\sigma = \frac{1}{1+\beta} > (<)\frac{1}{1-\mu}$ implies that $\mu < (>) - \beta$.

$$\bar{y}(K) = (aK^{-\beta} + b\overline{L}^{-\beta})^{\gamma} - (aK^{-\beta} + b\underline{L}^{-\beta})^{\gamma}$$

where $\underline{L}$ is the minimum labor and $\overline{L}$ is the maximum labor. The derivative of the range of output with respect to the capital stock is given by

$$\frac{\partial \bar{y}}{\partial K} = -a\gamma\beta K^{-(\beta+1)}\left((aK^{-\beta} + b\overline{L}^{-\beta})^{\gamma-1} - (aK^{-\beta} + b\underline{L}^{-\beta})^{\gamma-1}\right)$$

When $\gamma < (>)1$, $\frac{\partial \bar{y}}{\partial K} < (>)0$. This in turn implies that when $\gamma < (>)1$ then the level of capital that gives the maximum range of output, $\hat{K}$, is equal to the minimum (maximum) stock of capital.

□

With the most flexible level of capital so defined, and our definition of the irreversibility effect, consider whether or not the irreversibility effect holds in this example of a firm's demand for capital. According to our definition the irreversibility effect holds if $|K^* - \hat{K}| \leq |K^{**} - \hat{K}|$, where K^* is the level of capital investment in the first period with learning, and K^{**} the level with less or no learning. When $\sigma > \frac{1}{1-\mu}$ or $\gamma < 1$, then $\hat{K} = \underline{K}$ and the irreversibility effect holds if $|K^* - \underline{K}| \leq |K^{**} - \underline{K}|$, or when $K^* \leq K^{**}$. Since demand for capital declines with learning under these model parameters, the irreversibility effect

[6]If $\sigma = 1$ so that the production function is a Cobb-Douglas then one can show that a higher level of capital gives a greater range of output. If $\sigma = 0$ so that the production function is a Leontief then it is difficult to determine what level of capital gives greater flexibility tomorrow.

holds. Similarly, it can be established that when $\sigma < \frac{1}{1-\mu}$, the irreversibility effect holds if $K^* \geq K^{**}$, and that this is in fact the case.

4. Necessary Versus Sufficient Conditions

Building on the sufficient conditions established by Epstein, the literature has established other sufficient conditions for the irreversibility effect to hold, but only one necessary, and therefore more restrictive, condition, namely the condition established by Freixas & Laffont (1984) for intertemporally separable net benefit functions. With such functions, and using a sharper irreversibility constraint, $x_2 \leq x_1$, the dynamic optimization problem discussed in Section 2 simplifies to the following:

$$\max_{x_1 \in C_1} \left(B_1(x_1) + \sum_j q_j \max_{x_2 \leq x_1} \sum_i \pi_{ij} B_2(x_2, z_i) \right). \tag{8}$$

Since x_2 is constrained to be less than or equal to x_1, a larger value of x_1 in the first period gives the decisionmaker more flexibility in the second period. And therefore the irreversibility effect is said to hold so long as $x_1^* \geq x_1^{**}$. Theorem 2 specifies the condition developed by Freixas and Laffont for the irreversibility effect to hold given that a unique solution exists.

Theorem 2. *$x_1^* \geq x_1^{**}$ if $B_1(x_1) + \sum_j q_j J(x_1, \xi)$ is quasi-concave.*

The theorem states that the irreversibility effect holds if the value function is quasi-concave. Freixas and Laffont establish sufficiency analytically and then develop a numerical example to show that the sufficient condition is also necessary. Their numerical example establishes that if quasi-concavity is violated then, in fact, the irreversibility effect is violated. The irreversibility effect is also shown to be equivalent to $J(x_1, \xi') - J(x_1, \xi)$ being locally increasing in x_1 where ξ' is the more informative signal, ξ is the less informative signal, and $J(x_1, \xi') - J(x_1, \xi)$ is the value of information.

A slight modification, however, of Frexias and Laffont's numerical example shows that quasi-concavity is only sufficient. In Freixas and Laffont's numerical example the random

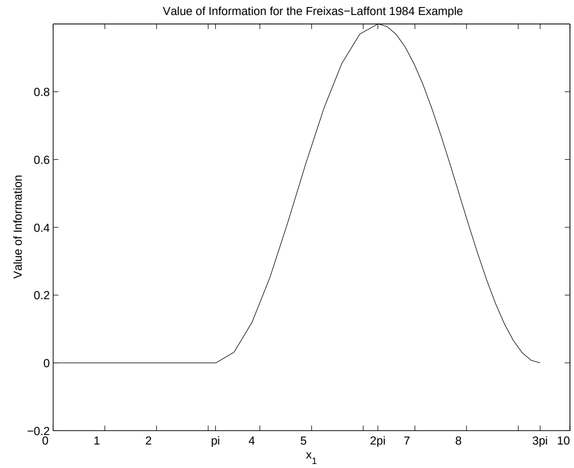

FIGURE 1. Value of Information for the Example in Freixas-Laffont 1984

variable z is assumed to take two possible values, z_1 and z_2, each with probability 0.5. Furthermore, there are two levels of learning, perfect or none at all. The functional forms of the benefit functions are,

$$B_1(x_1, 0 \leq x_1 \leq 2.5\pi) = \pi$$

$$B_1(x_1, x_1 \geq 2.5\pi) = -1.25(x_1 - 2.5\pi) + \pi$$

$$B_2(x_2, z_1) = 2x_2$$

$$B_2(x_2, z_2) = -\cos x_2 + 1$$

With these benefit functions quasi-concavity is violated, $x_1^* \leq x_1^{**}$ and $J(x_1, \xi') - J(x_1, \xi)$, the value of information, is not increasing in x_1. These results are shown in Figure 1 where the choice variable in the first period, x_1, is drawn on the x-axis and the value of information on the y-axis. Note that in the range $[2\pi, 3\pi]$ the value of information decreases in x_1. So long as the optima lie in this range the irreversibility effect is violated.

Now consider a slight modification where $B_1(x_1)$ and $B_2(x_2, z_1)$ remain unchanged and $B_2(x_2, z_2)$ is given by

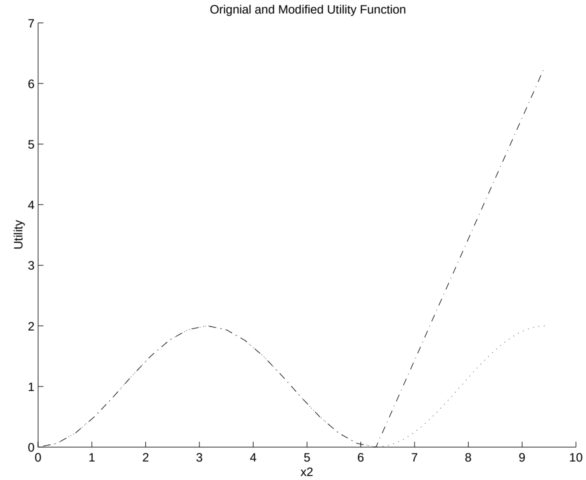

FIGURE 2. Original and Modified Utility Functions

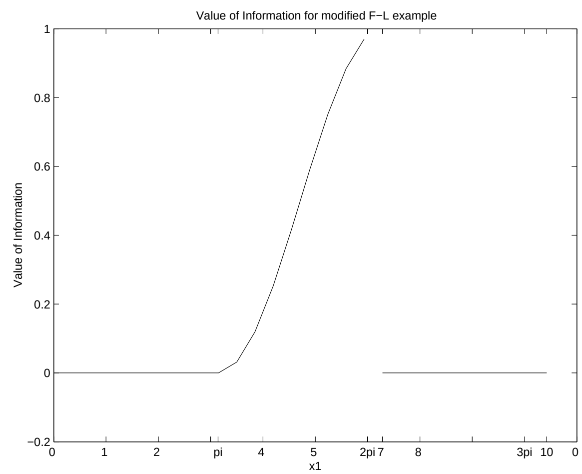

FIGURE 3. Value of Information for the Modified Example in Freixas-Laffont 1984

$$B_2(x_2, z_2, 0 \leq x_2 \leq 2\pi) = -\cos x_2 + 1$$

$$B_2(x_2, z_2, x_2 \geq 2\pi) = x_2 - 2\pi$$

The original and modified benefit functions are illustrated in Figure 2. Both functions are identical in the range $[0, 2\pi]$ and thereafter the original function is represented by the dotted line and the modified function by the broken line.

With the modified value function although quasi-concavity is still violated, $x_1^* \geq x_1^{**}$ (strictly greater if the optimal lies between π and 2π and equal otherwise) and the value of information is no longer decreasing over a finite interval of x_1. These results are shown in Figure 3. The irreversibility effect holds though quasi-concavity is violated. Consequently, quasi-concavity is not necessary, merely sufficient, for the irreversibility effect to hold in the class of intertemporally separable benefit functions. We note however that quasi-concavity is a very weak condition. Most, if not all, empirically important benefit functions will exhibit this property.

In summary, to date a number of sufficient conditions have been established for the irreversibility effect, but not necessary conditions, which would be both more restrictive and more powerful.

5. Intertemporally Non-separable Benefits

An issue of some importance in the literature is whether or not the benefit function is separable in x_1 and x_2. In fact, the existing sufficient conditions in the literature can be organized into two broad categories: those that apply to models with separable benefit functions and those that apply to models with non separable functions. Conditions developed by Freixas & Laffont (1984) and Kolstad (1996), for example, apply to separable models, while those developed by Epstein (1980), Ulph & Ulph (1997), and Gollier et al. (2000), to non-separable models.

There is, however, a perception in the literature that Epstein's condition can not be used to investigate the irreversibility effect in some models with intertemporally non separable benefit functions, in particular models of global warming where damages depend on the amount of accumulated greenhouse gases making the damage function non separable (for example, see Ulph & Ulph (1997) and Gollier et al. (2000)). As shown in Section 3, so long as one uses our more general definition of the irreversibility effect, which, in particular, forces one to determine the point of maximum flexibility, Epstein's sufficient condition can be used for at least some intertemporally non separable benefit functions. In this section, we establish that Epstein's conditions can be used in the environmental application to the optimal control of greenhouse gas emissions. This also helps us to identify a class of models where in fact the irreversibility effect is violated.

Consider a simple variant of the global warming model presented by Ulph & Ulph (1997) where the decisionmaker solves the following dynamic optimization problem:[7]

$$\max_{0 \le \delta e_0 \le x} \left(V(x - \delta e_0) + \rho \sum_j q_j \max_{0 \le \delta x \le y} \big(W(y - \delta x) - \sum_i \pi_{ij} \theta_i D(y) \big) \right), \tag{9}$$

where $V(.)$ is the first period benefit function, $W(.)$ the benefit function in the second period, and $D(.)$ the damage function. Furthermore, e_0 is the initial stock of greenhouse gases, x the stock at the end of the first period, and y the stock of greenhouse gases at the end of the second period. $(1 - \delta)$ is the rate of decay of the stock of greenhouse gases, ρ the discount factor, and θ a discrete random variable that reflects the underlying uncertainty about the potential damages from the stock of greenhouse gases. q_j and π_{ij} are as previously defined in Section 2. The flows of greenhouse gases in both periods are constrained to be non-negative, which, in turn, imply that the stock of greenhouse gases at the end of the second period is constrained to be no smaller than δ times the stock at the end of the first period. With these assumptions a smaller stock at the end of the first period gives the decision maker greater flexibility. Therefore, $\hat{x} = \delta e_0$ and the irreversibility effect then holds if $|x^* - \delta e_0| \le |x^{**} - \delta e_0|$, or if $x^* \le x^{**}$.

This problem is structurally identical to the consumption and savings problem discussed by Epstein (1980) in which an individual allocates an initial amount of wealth between consumption and savings over three periods. Investment in the first period yields a fixed return while investment in the second period yields a random return. Some information is gained about the random rate of return at the beginning of the second period. The individual solves the following problem:

[7] This problem is a slight generalization of Ulph and Ulph in that it allows for a range of learning levels while Ulph and Ulph allow for either perfect or no learning.

$$\max_{0 \le x_1 \le w_1} \left(U(w_1 - x_1) + \frac{1}{\beta} \sum_j q_j \max_{0 \le x_2 \le r x_1} \big(U(r x_1 - x_2) + \frac{1}{\beta} \sum_i \pi_{ij} U(x_2 z_i) \big) \right), \tag{10}$$

where x_1 and x_2 denote savings in periods 1 and 2 respectively, w_1 the initial wealth, $\frac{1}{\beta}$ is the discount factor, r the sure gross rate of return to first period savings realized at the beginning of period 2, and z_i the random gross return to second period savings that is realized at the beginning of period 3.[8] Since savings in the second period are constrained to be less than r times the savings at the end of the first period, higher savings in the first period lead to greater flexibility in the second period. Therefore, $\hat{x}_1 = w$ and the irreversibility effect is said to hold if $|x_1^* - w| \le |x_1^{**} - w|$. Since consumption is constrained to be positive, $x_1 < w$, and the condition simplifies to $x_1^* \ge x_1^{**}$.

Using the following constant relative risk aversion utility function,

$$B(c) = \begin{cases} \frac{c^{1-\alpha}}{1-\alpha} & \text{if } \alpha \neq 1, \\ \log(c) & \text{if } \alpha = 1, \end{cases} \tag{11}$$

where α is the coefficient of relative risk aversion (CRRA), Epstein establishes that the effect of learning on the optimal level of savings in the first period depends on the elasticity of intertemporal substitution, that is, on $\sigma = \frac{1}{\alpha}$. When $\sigma > 1$ the slope of the value function is convex and the possibility of learning about the future rate of return leads to an increase in savings in the first period and when $\sigma < 1$ the possibility of learning leads to a decrease in the level of first period savings.[9] Since the level of first period savings does not unambiguously

[8] There are two minor differences between the models. First, the consumption-savings problem spans three periods while the global warming problem spans two. Second, the global warming problem assumes a multiplicative form for the underlying uncertainty while the consumptions-savings model places no such restriction on the nature of uncertainty. In terms of the model described in Section 2, multiplicative uncertainty implies that $B_2(x_1, x_2, z_i) = z_i B_2(x_1, x_2)$. The first difference is of no consequence as there is no additional learning in the three-period model while the second difference makes the global warming problem a special case of the consumption-savings problem.

[9] Note that Gollier et al. (2000) generalize this result by showing that it can be extended to a more general class of preferences, hyperbolic absolute risk aversion. Specifically, within the class of models characterized

increase with learning, this is evidence that the irreversibility effect is sometimes violated in this problem. Specifically, the irreversibility effect is violated when $\sigma < 1$, that is, when benefits are intertemporally non-substitutable or the coefficient of relative risk aversion is large. In fact, and as established by Ha-Duong & Treich (2004) using isoelastic preferences that, unlike Von-Neumann-Morgenstern perferences, do not constrain risk aversion to be the reciprocal of the coefficient of risk aversion, the violation of the irreversibility effect in this case is caused by a low elasticity of intertemporal substitution and not by a high coefficient of relative risk aversion.

Given the structural similarity between this and the global warming problem, this result implies that the irreversibility effect will be violated under certain model parameters for the global warming problem. Specifically, following Epstein's proof, it is easy to show that when $\sigma < 1$ the slope of the value function for the global warming problem is convex, and an increase in learning leads to an increase in the stock of greenhouse gases in the first period. The irreversibility effect is therefore violated when $\sigma < 1$.

Ulph and Ulph, however, do not use constant relative risk aversion preferences but instead assume that the benefit and the damage functions are quadratic. Specifically, $V(x - \delta e_0) = a_1(x - \delta e_0) - 0.5a_2(x - \delta e_0)^2$, $W(y - \delta x) = a_1(y - \delta x) - 0.5a_2(y - \delta x)^2$ and $D(y) = 0.5a_3y^2$. With these preferences, the coefficient of relative risk aversion associated with the net benefit function in the second period,

$$CRRA = \frac{(a_2 + a_3)y}{a_1 + a_2\delta x - (a_2 + a_3)y}.$$

Furthermore, it is easy to show that $CRRA \gtreqless 1$ iff $2(a_2 + a_3)y \gtreqless (a_1 + a_2\delta x)$. Whether CRRA is greater or less than one then depends on parameter values and on the values

by hyperbolic absolute risk aversion preferences, that is, with utility functions

$$B(x) = \frac{\gamma}{1-\gamma}\left[\eta + \frac{x}{\gamma}\right]^{1-\gamma}, \qquad (12)$$

where x is a function of x_1 and x_2, and the coefficient of absolute risk aversion is $\eta + \frac{x}{\gamma}$, the slope of the value function is concave (convex) in the random variable if and only if $\gamma < 1$ ($\gamma > 1$ or $\gamma < 0$). Note that if $\eta = 0$ in equation (12), then hyperbolic absolute risk preferences reduce to constant relative risk aversion preferences and γ can be interpreted as the coefficient of relative risk aversion.

of endogenous variables, namely, the stock of greenhouse gases in the first and the second period. Ulph and Ulph also establish that with these preferences the slope of the value function is neither concave nor convex, and conclude that Epstein's sufficient condition cannot be applied to a model of global warming with non separable benefit functions. The authors accordingly develop a new sufficient condition: the irreversibility effect is said to hold if the irreversibility constraint bites when there is no possibility of learning. In terms of our canonical model, let x_2^{**} and x_1^{**} denote the optimal decisions in the absence of learning. If $x_1 \leq x_2$, then $x_2^{**} = x_1^{**}$ implies that the irreversibility constraint bites in the absence of learning.[10] Note, however, that it is the functional form of the benefit and damage functions that yields the ambiguity, and causes Epstein's condition to fail, and not the fact that the global warming problem is inherently nonseparable.

[11]

6. Concluding Remarks

In this paper we set out to interpret and extend the strand of the irreversibility effect literature that has focused on the question of whether the rather strong and unambiguous results of Arrow and Fisher, Henry, and Dixit and Pindyck, continue to hold in still more general settings in which the intertemporal benefit function exhibits properties not considered by these authors. We have tried to build on this literature in a number of ways.

We have introduced a more general definition of the irreversibility effect. Our definition encompasses the existing ones, and is independent of the structure of the problem. It applies to problems where the existing definitions fail, namely where the decisionmaker chooses two

[10]In the same vein Kolstad (1996) shows that the irreversibility effect holds in models with *effective* irreversibility—that is, in models in which the irreversibility constraint bites.

[11]As pointed out by one of our reviewers, there are two types of ambiguities that arise in the application of Epstein's conditions. The first, and the one that arises in the firm's-demand-for-capital example, and in the consumption-savings example, arises because the slope of the value function varies with model parameters. *A priori*, and without specifying the values of model parameters, one cannot determine whether the slope of the value function is concave or convex. However, once the parameter values have been specified, one can determine unambiguously whether the slope is convex or concave. The second type of ambiguity, the one referred to by Ulph & Ulph (1997), arises when the slope of the first derivative of the value function is neither concave nor convex irrespective of the parameter values. That is, even once the parameter values have been specified, one cannot determine whether the slope of the value function is concave or convex.

different objects in the first and the second periods, say producible capital or protected land in the first and the relevant labor complement in the second. In such problems the existing definitions are not helpful. Our definition, on the other hand, by forcing one to consider which choice generates the most flexibility, continues to apply.

Applications of our definition, and a numerical example which shows that the only necessary condition in the literature is only sufficient, establish that the irreversibility effect holds more widely than has perhaps previously been recognized. Another interesting interpretive result is that Epstein's condition, the original contribution to this literature, and Theorem 1 in our paper, can in fact be applied more widely, in particular to intertemporally nonseparable benefit functions, as in the global warming problem, than previously indicated.

References

Albers, H. J. & Goldbach, M. J. (2000). Irreversible ecosystem change, species competition, and shifting cultivation, *Land Economics* **22**.

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

de Angelis, H. & Skvarca, P. (2003). Glacier surge after ice shelf collapse, *Science* **299**.

Dixit, A. K. & Pindyck, R. S. (1994). *Investment under Uncertainty*, Princeton University Press.

Epstein, L. G. (1980). Decision making and the temporal resolution of uncertainty, *International Economic Review* **21**: 269–283.

Freixas, X. & Laffont, J.-J. (1984). On the irreversibility effect, *in* M. Boyer & R. Kihlstrom (eds), *Bayesian Models in Economic Theory*, NHPC, pp. 105–114.

Gollier, C., Jullien, B. & Treich, N. (2000). Scientific progress and irreversibility: an economic interpretation of the precautionary principle, *Journal of Public Economics* **75**: 229–253.

Ha-Duong, M. & Treich, N. (2004). Risk aversion, intergenerational equity and climate change, *Environmental and Resource Economics* **28**.

Hanemann, W. M. (1989). Information and the concept of option value, *Journal of Environmental Economics and Management* **16**: 23–37.

Hartman, R. (1976). Factor demand with output price uncertainty, *American Economic Review* **66**: 675–681.

Henry, C. (1974). Investment decisions under uncertainty: The irreversibility effect, *American Economic Review* **64**: 1006–12.

Hirshleifer, J. & Riley, J. (1992). *The Analytics of Uncertainty and Information*, Cambridge University Press.

Jones, R. A. & Ostroy, J. M. (1984). Flexibility and uncertainty, *Review of Economic Studies* **51**: 13–32.

Kerr, R. A. (1998). West antarctica's weak underbelly giving way ?, *Science* **281**: 499–500.

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

Marschak, J. & Miyasawa, K. (1968). Economic comparability of information systems, *International Economic Review* **9**: 137–174.

Schultz, P. A. & Kasting, J. F. (1997). Optimal reductions in co_2 emissions, *Energy Policy* **25**: 491–500.

Ulph, A. & Ulph, D. (1997). Global warming, irreversibility and learning, *The Economic Journal* **107**: 636–650.