

# Individual Utility in a Context of Asymmetric Sensitivity to Pleasure and Pain: An Interpretation of Bentham's Felicific Calculus

André Lapidus, Nathalie Sigot

► **To cite this version:**

André Lapidus, Nathalie Sigot. Individual Utility in a Context of Asymmetric Sensitivity to Pleasure and Pain: An Interpretation of Bentham's Felicific Calculus. *European Journal of the History of Economic Thought*, Taylor

Francis (Routledge), 2000, 7 (1), pp.45-78. <hal-00344899>

**HAL Id: hal-00344899**

**<https://hal-paris1.archives-ouvertes.fr/hal-00344899>**

Submitted on 6 Dec 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**INDIVIDUAL UTILITY  
IN A CONTEXT OF ASYMETRIC SENSITIVITY  
TO PLEASURE AND PAIN :  
AN INTERPRETATION OF BENTHAM'S *FELICIFIC CALCULUS***

André Lapidus\* / Nathalie Sigot\*\*

*European Journal of the History of Economic Thought*

7(1) Spring 2000, pp. 45-78

“Coupe les arbres, si tu veux, casse aussi les pierres  
mais prends garde,  
prends garde à la lumière livide de l'utilité”

(André Breton, Paul Eluard, *L'Immaculée Conception*)

## **0. Introduction: after Jevons, who needs Bentham ?**

It is generally agreed that the developments that Jeremy Bentham devoted to the positive aspects of the “principle of utility” exerted some influence over the rise of modern analysis of individual behaviour (see, for example, Stark 1946; Schumpeter 1954 or, more recently, Black 1988). Such an opinion would be confirmed by Jevons' acknowledgement of his being indebted to the Benthamite calculus of pleasure and pain (1871:27<sup>1</sup>). Meanwhile, Jevons' decision to root his own contribution in Bentham's work rather than in more recent accounts of utilitarianism – such as John Stuart Mill's – is far from being some kind of ritual tribute, paid to the founder of the doctrine: Bentham was invoked in order both to testify the break with British political economy of the second half of the nineteenth century, and to give this break the legitimacy of an ancient tradition.

---

\* University of Paris I Panthéon-Sorbonne, CHPE. Address for correspondence: CHPE, Maison des Sciences Economiques, Université de Paris 1 Panthéon-Sorbonne, 106 Bd de l'Hôpital, 75647 Paris Cedex 13, France. E-mail: lapidus@univ-paris1.fr

\*\* University of Besançon, CHPE and ERSI. Address for correspondence: CHPE, Maison des Sciences Economiques, Université de Paris 1 Panthéon-Sorbonne, 106 Bd de l'Hôpital, 75647 Paris Cedex 13, France. E-mail: nsigot@univ-paris1.fr.

<sup>1</sup> “The theory which follows is entirely based on a calculus of pleasure and pain”; see also Jevons, 1862: § 2-7.

It could be argued, then, that Jevons, among others, inherited some parts of the Benthamite legacy. Nonetheless, this does not contradict the idea that some essential features of Bentham's felicific calculus had been given up in the analysis of utility introduced in the *Theory of Political Economy*. Therefore, this paper precisely aims at restituting these aspects of the Benthamite scope, cancelled out by Jevons' contribution, but in no way inconsistent with contemporary approach to the analysis of utility.

To put it briefly, Jevons not only explicitly added differential calculus to some canonical representation of the individual which took shape since the mid eighteenth century; but he also locked up this individual into *given* tastes and sensitivity to pleasure and pain. Since Jevons, in other words, it has become usual to identify an economic agent by an exogenous order of preferences, or function of utility. Now, it must be recalled that such was not the case within Bentham's writings (see Sigot 1995: 26-74): although the Benthamite individual is unambiguously submitted to "these sovereign masters, *pain* and *pleasure*" (Bentham 1789: 1), he is, for this very reason, a changing individual. His preferences could no longer be considered as simply given, but as the outcome of the pleasure and pain that he previously enjoyed and suffered. More importantly, perhaps, than its use of the marginal principle, is the fact that Jevons' contribution surreptitiously led subsequent writers to desert the old intuition of endogenous preferences, in favour of the seemingly more fruitful idea of exogenous preferences.

In this paper, we attempt to restore Bentham's position in a more formal way – and in a more recent framework. Of course, this implies a restriction of the general principle, according to which tastes are submitted to permanent changes, to some simple analytical statement. The accepted changes therefore result from the individuals' ability to anticipate a change in their situation. Consequently, the agent's preference order will be viewed as depending on his initial situation, and on asymmetric sensitivity to gains and losses, relative to this situation (**section 1**). Bentham clearly expressed this idea when he argued that "the pleasure of gaining is not equal to the evil of losing" (1785-6: 331).

This leads to discuss the coexistence of multiple preference orders on the set of final situations, which might characterise a same individual (**section 2**). It will be shown that they are consistent, in the sense that they are associated to an "anticipated utility function", depending on an initial situation  $x(t_0)$ , and on an anticipated trajectory  $\{x(t)\}$ . Two important consequences proceed from the existence of an anticipated utility function.

Assuming that the agent's effectual choices follow his preferences, the first consequence is a representation of economic behaviour which allows what would to-day be considered as preference reversal phenomena (**section 3**), of which typical illustration is provided by Bentham's analysis of the optimal labour contract, and which could be extended to the analysis of demand and equilibrium.

The second consequence rests upon the ability to anticipate. Anticipation does not only concern the pleasure or pain generated by a move  $\{x(t)\}$  from  $x(t_0)$  to  $x(t_f)$ : it also concerns the new preference order resulting from this move. The agent is then faced with rival preference orders, associated with the moves he is able to achieve. Taking fully into account the possibility to anticipate, henceforth disconnects the question of preferences from that of choices. This can be interpreted as the introduction of a true deliberation, following the simple assessment of utility, into the theory of individual economic behaviour (**section 4**). Several features of early utilitarianism thus appear as closely linked: the systematic search for incentives is a response to a world in which the ability to anticipate fails among individual agents, and the emphasis on education is a way to improve this ability to anticipate. From an analytical point of view, this deliberation – from the legislator, or from the individual – aims then at achieving this non-conflicting order of preferences, on which Jevons was to build utility theory.

### ***1. De gustibus est disputandum***

One needs to be an economist trained in general equilibrium theory to assert without irony that an economic agent might be characterised by *given* preference order and set of choice, and that if preferences can indeed change, the reasons and the magnitude of this change are beyond our competence as economists. Otherwise, this is a rather schematical view. On the one hand, one could object, for instance, that Hayek (in the analysis of the mental process giving birth to a spontaneous order) or Duesenberry (1949) (with the relative income hypothesis) move away from the idea that exogeneous preferences generate economic behaviour. On the other hand, as D. Requier-Desjardins (1996) noticed it, such contributions like C.C. Von Weizsäcker (1971) (which stresses the difficulty to build short-term indifference curves, since the individual is insufficiently informed on his preferences on the long run) shifts from the traditional view, in the sense that it focuses on the necessity to identify the law of evolution of individual preferences. But Becker's and Stigler's 1977 attempt to reduce endogeneous to exogeneous preferences illustrate the fact that, more than a powerful instrument for economic calculus, the latter are also a flexible concept, able to grasp what is usually considered as a change in tastes<sup>2</sup>. In contrast, it is a common sense opinion that our tastes might change just as we are, ourselves, changing, and that these changes are worth being examined by economists.

---

<sup>2</sup> Becker and Stigler (1977) favoured several ways of reducing seemingly endogeneous to exogeneous preferences. For instance, 1) a change in shadow prices rather than in tastes (an "unilluminating 'explanation'", said Becker and Stigler, 1977 : 78), illustrated by A. Marshall's example of good music generating an increasing demand for good music; 2) additional information, such as transmitted by publicity, which gives birth to a consistent change in choices – not in tastes.

Our primary intuition is that such a common sense opinion was, in some way, as trite for most scholars working on economic matters before the end of the nineteenth century, as it is now. Of course, Bentham was one of those scholars. But, whereas the assumption that preferences are exogenous to individual is nowadays explicit, the reverse assumption – endogenous preferences – hardly supplies direct evidence in Bentham’s writings. Nevertheless, indirect evidence (§ 1.1. and 1.2.) seems conclusive. This will lead us to focus on a principle recalled by Bentham on several occasions, and to which too little attention has been previously paid: “pleasure of gaining is not equal to the evil of losing” (§ 1.3.).

### 1.1. An interpretation of the taxonomy of pleasures and pains

Chapter IV of Bentham’s *Introduction to the Principles of Morals and Legislation* [1789] does not seriously challenge the current conception of an economic agent ruled by a given order of preferences. Dedicated to the identification of the dimensions of pleasure and pain<sup>3</sup> and to the resulting measure of happiness, these pages constituted the first step in the constitution of Jevon’s value theory (1862: 282-3 and 1871: 27 and 33-4), but this latter seems to have been wise enough to abandon promptly Bentham’s pioneering perspective: trying to measure utility through that which causes it, might still be of some interest for a psychologist; it is a long time since it is only of poor relevance for an economist.

But chapter V’s obsessional taxonomy of pleasures and pains is far more puzzling, since the fourteen elementary pleasures and the twelve elementary pains which Bentham carefully distinguishes<sup>4</sup> help the reader understand, not that the value attributed to a thing is connected to the happiness it provides, but the reasons of this happiness, in terms of the natures of the pleasures and pains involved.

---

<sup>3</sup> Bentham isolated four criteria, namely: “intensity”, “duration”, “certainty or uncertainty”, “propinquity or remoteness”, to which he added “fecundity”, purity” and “extent” (1789: 29-30).

<sup>4</sup> The simple pleasures which Bentham enumerates run as follows: “1. The pleasure of sense. 2 The pleasures of wealth. 3. The pleasures of skill. 4. The pleasures of amity. 5. The pleasures of a good name. 6. The pleasure of power. 7. The pleasures piety. 8. The pleasures of benevolence. 9. The pleasures of malevolence. 10. The pleasures of memory. 11. The pleasures of imagination. 12. The pleasures of expectation. 13. The pleasures dependent on association. 14. The pleasures of relief”. The pains are those of “1. [...] privation. 2. The pains of the senses. 3. The pains of awkwardness. 4. The pains of enmity. 5. The pains of an ill name. 6. The pains of piety. 7. The pains of benevolence. 8. The pains of malevolence. 9. The pains of the memory. 10. The pains of the imagination. 11. The pains of expectation. 12. The pains dependent on association” (1789: 33-4). Bentham did not find necessary to give any justification for this “catalogue”. He just noticed that this was “what seemed to be a complete list of the several simple pleasures and pains of which human nature is susceptible”, adding that “[i]t might perhaps have been a satisfaction to the reader, to have seen an analytical view of the subject, taken upon an exhaustive plan, for the purpose of demonstrating the catalogue to be what he it purports to be, a complete one. The catalogue is in fact the result of such an analysis; which, however, I thought it better to discard at present, as being of too metaphysical a cast, and not within the limits of this design” (*Ibid.*: 34 n. 1).

These reasons are not simply extra-economical considerations. Whereas Bentham's chapter IV is consistent with the contemporary idea of an economic agent characterised by exogenous preferences, chapter V does not manifest the same consistency. The picture of the individual drawn by Bentham is not only a far wider picture than the one commonly associated to an economic agent: it also implies a different picture of this economic agent.

More precisely, the pleasure of the senses made up from the pleasure of novelty, the pleasure of acquisition, as an instance of the pleasure of wealth, the pleasures of memory, and those of relief, the pains of privation or of senses frustration, of the memory, of imagination or of expectation (1789: 33-42), could hardly be imagined within the framework of an individual whose tastes and sensitivity to pleasures and pains remain unchanged. As soon as we investigate not only the dimensions and measure of pleasures and pains, but also the content of the latter, we are urged to acknowledge that the actualisation of such pleasure or such pain will change the preferences of the one who enjoyed or suffered it. This interpretation is confirmed by Bentham's chapter VI, in which he tries to isolate the thirty-two circumstances (health, strength, knowledge, etc...) which might affect the influence on happiness of such pleasure or pain. Obviously, whereas age, government or – to a certain extent – health, seem to be external circumstances, most of the latter are not strictly external but, at least partly (like religion, profession, or habitual occupation), a consequence of the individual's previous choices.

This complex drawing of a changing individual should not be a surprising one within Bentham's work. It is linked both to the psychological foundations of his interpretation of the principle of utility, and to his social project.

## **1.2. Associationism at the root of individual changes**

The connection between the principle of utility and associationist psychology, inspired, at Bentham's time, by the works of Hume and Hartley (Bentham, 1789: 43-69), deserves special attention. This connection implies that the conclusions of the *felicific calculus* depend, for each individual, on a continuous interaction between feelings, ideas, and sentiments. But this interaction is also a source of permanent transformations of the individual, of renewed receptivity to the different events which originate these major feelings – pleasure and pain.

This possibility of transformation, opened by associationist psychology, therefore constitutes a prerequisite of the numerous – and generally sterile<sup>5</sup> – projects that Bentham

---

<sup>5</sup> This commentary must be qualified. It is usually acknowledged that Bentham's influence on penal law is far from negligible. In England, for instance, this influence was both direct and indirect, through some of his followers, like John Austin and Henry Brougham (E. Halevy, 1901-4). According to J. Pradel (1989: 49-50),

elaborated all along his intellectual life. As a social reformer, he never imagined that institutional changes would only modify the set of choices of each individual, leaving unchanged his preferences. When he wrote on such topics as the penitential system, the management of the poor, the liberty of the press, or the education, his objective was to change human beings, and not only institutions – i.e. preferences, and not only sets of choices. It is worth noticing that such a conception of social reforms was not a real issue of the debates in which Bentham was involved: after all, even his opponents shared this simple and quite common idea that changes in institutions might generate changes in human beings<sup>6</sup>. However, the link with associationism was a typical feature of his position. At first glance, one could imagine that this link was rather loose in early writings. But it did become explicit after Bentham's meeting with James Mill in 1808, when collaboration between the two men gave birth to what Elie Halévy called "philosophical radicalism".

A significant example of the part played by associationist psychology can be found in Bentham's increasing interest in education<sup>7</sup>. Some ten years before his meeting with J. Mill, Bentham took up the theme of education as a particular side of the Poor question, which held all his attention (1797-8). But in 1816, when he published *Chrestomathia*, education was shown to be of the utmost importance, to such an extent that his commitment in favour of the "Lancasterian system" induced him to offer a part of his own garden, so as to edificate there a "Chresthomatic School". As usual, the project collapsed, and only the creation of the University College of London, in 1830, has kept trail of the initial ambition of its author. Accordingly, although J. Mill never made Bentham discover an unexplored field, he certainly helped him strengthen an argument.

The foundations of J. Mill's analysis of education are provided in the article "Education", written between 1818 and 1819 and published in the *Encyclopaedia Britannica* and in his *Analysis of the Phenomena of the Human Mind* [1829]. For J. Mill, human behaviour depends on sequences of ideas which are copies of sensations. "The sensations which we have through the medium of the senses", wrote J. Mill, "exist only by the presence of the object and cease upon its absence [...]. It is known part of our constitution, that when our sensations cease, by the absence of their objects, something remains [...]. We have two classes of feelings; one that which exists when the object of sense is present; another, that which exists after the object of sense has ceased to be present. The one class of feelings I call SENSATIONS; the other class of feelings I call IDEAS" (1829: 51-2). The connection between

---

Bentham's stamp is also evident on the French penal code of 1810, on the Louisiana code of 1821, or on the Indian code of 1833.

<sup>6</sup> See, for instance, Sir G.O. Paul's indictment of Bentham's Panopticon, which reflected the opinion of the Parliamentary Commission which, reexamining the project in 1810, jeopardized its realisation (Semple 1993: 268 and Hume 1974: 48).

<sup>7</sup> On the question of education in early utilitarianism, see Sigot 1995: 271-87.

sensations and ideas is governed, according to J. Mill, by two rules of association (*ibid.*: 71): the “synchronous order” – which Hume called “space contiguity” in the *Treatise on Human Nature* – and the “successive order” – which refers to the Humian principles of “contiguity in time” and “causation”<sup>8</sup>. And the strength of this connection will, in return, depend on the strength of the sensations, and on the frequency of their association. Here lies the theoretical basis of the educational project: modifying human behaviour means either acting upon the sequences of ideas or creating new ones. Education hence consists in implementing “virtuous” sensations of different magnitudes and, above all, in increasing the frequency of their association (1818-9: 147 and 151).

### **1.3. The pleasure of gaining is not equal to the evil of losing**

No doubt, such a complex picture of the determinants of individual behaviour limited drastically the part of the Benthamite legacy that Jevons was ready to make his own: founding economic calculus on the basis of a given utility function was already a difficult task, which required nearly a century after Jevons to be achieved; but the enterprise would surely have been bound to fail with a utility function submitted to continuous changes.

However, some kinds of changes, playing a crucial part in Bentham’s analysis, could nowadays be dealt with in the framework of usual utility theory. They appear in Bentham’s theory as the consequence of a general principle concerning the respective effects of gains and losses on individual happiness. While discussing the efficiency of legal institutions, Bentham introduced it as follows: “pleasure of gaining is not equal to the evil of losing” (1785-6: 331). As such, the legislator was expected to take it into account when instituting incentives aiming at modifying individual behaviour. Over half a century, Bentham expressed the same idea in different contexts: in the “Preface” to the *Introduction to the Principles of Morals and Legislation* (p. xxv, n. 1), in *Pannomial Fragments* (p. 229), or in the manuscript *Sur les Prix* (f. 48). Although Bentham did succeed to systematise the consequences of this statement on the art of the legislator, the idea of an asymmetry between pain and pleasure, granting more weight to the former at the expense of the latter, was far from being pathbreaking when it appeared under his pen. As M. Guidi (1993) noticed it, its origin could be traced back to Locke, Verri or Maupertuis<sup>9</sup>. Anyway, this principle was to Bentham

---

<sup>8</sup> For J. Mill, Hume’s causation could be reduced to contiguity in time (1829: 110).

<sup>9</sup> However, M. Guidi (1993: 51-8) advocated the thesis of a break within Bentham’s work, which gave birth to a “perfect symmetry” between pain and pleasure. This interpretation needs to be qualified. It seems that as Bentham went along his work, he became more and more aware of the negative effects, on general happiness, of individual asymmetric sensitivity to pain and pleasure (see *infra* § 3.3). As a result, he seems to have established a distinction between i) the positive dimension of the principle of utility, from an individual point of view, according to which asymmetric sensitivity to pain and pleasure keeps on with playing a decisive part in individual behaviour, and ii) its normative dimension, from the point of view of the legislator, which rests on



universal enough to be called an “axiom”, and unreasonable enough to be viewed as an “axiom of Mental Pathology”.

A primary intuition of the meaning of this principle would be that it just states that marginal utility is decreasing. But it is not necessary to make such a strong assumption as the concavity of the utility function to give sense to the principle “pleasure of gaining is not equal to the evil of losing”. Let us assume that the agent’s initial situation is represented by  $x^0 = (x_1^0, x_2^0)$ . When he anticipates a variation  $dx$  of his situation, he also anticipates a variation  $dU$  of his level of utility. Imagine then two opposite anticipated variations of the amount of good 1,  $dx_1$  and  $-dx_1$  (where  $dx_1 \geq 0$ ), leading to a same anticipated variation of utility  $dU$ . Asserting that the “pleasure of gaining is not equal to the evil of losing”, therefore means that  $dx_2$  will be greater, relatively to  $dx_1$ , when it has to compensate a loss  $-dx_1$  than when it has to compensate a gain  $dx_1$ . Of course, a similar conclusion could be obtained, assuming anticipated variations  $dx_2$  and  $-dx_2$ . This clearly means that, if a function of anticipated utility exists, it is not continuously differentiable. In  $x^0$ , its left-hand first derivatives,  $u_1^-$  and  $u_2^-$ , should be respectively greater than its right-hand derivatives,  $u_1^+$  and  $u_2^+$ , so that the marginal rates of substitution might be ordered as follows:

$$u_1^-/u_2^+ > u_1^+/u_2^+ > u_1^+/u_2^- \quad [1.1]$$

$$u_1^-/u_2^+ > u_1^-/u_2^- > u_1^+/u_2^- \quad [1.2]$$

If we remain in the neighbourhood of  $x^0$ , this representation is not so far removed from the traditional one: it also allows the function of anticipated utility to be locally increasing and strictly quasi-concave. The only noticeable difference is that in  $x^0$ , the marginal rate of substitution depends on the respective signs of the anticipated moves  $dx_1$  and  $dx_2$ .

Up to this point, Bentham’s principle might seem a little bit “pathological”, but it does not bear expressly preferences changes. In order to obtain this, we need to take more completely into account the ability to anticipate that it presupposes. The above mentioned  $dx$  moves are *anticipated* moves. The resulting  $dU$  are *anticipated* variations of utility. Therefore, if an individual is able to anticipate the consequences of a move, in terms of utility, he is also able to anticipate a change in his initial situation. But, as he anticipates such a change, he also anticipates a move of the discontinuity of the marginal rate of substitution and, consequently, a change in his local preferences.

More generally, the idea that “pleasure of gaining is not equal to the evil of losing” implies that sensitivity to gain and loss will change along with the initial situation. In this respect, this seemingly simple statement of Bentham allows, alone, preferences over final

---

some kind of symmetry between pain and pleasure - a symmetry which progressively forced itself upon individual agents, through deliberation and education (see *infra* § 4).

situations to be not given exogenously, but relative to the agent's initial situation. In what follows, we will explore some of its analytical consequences.

## 2. A representation of “Benthamite” individual utility

Asymmetric sensitivity to gains and losses would not to-day imply the complete rejection of the analytical apparatus accumulated since Jevons. But it surely induces some revision. In particular, utility would appear as depending on a trajectory mentally followed by the agent (§ 2.1.). On first view, such a representation seems rather puzzling, since it opens the path to the possibility that a same final situation of alternative trajectories is associated with different levels of utility. This gives rise to the identification of optimal trajectories, the final situations of which constitute maps of indifference relative to each initial situation (§ 2.2.).

### 2.1. Sensitivity to gains and losses and the anticipated utility function

In traditional theory of utility, the effects of a move  $dx$  from any initial situation  $x$  can be appreciated through the use of an invariant, the utility function. The invariant, here, will be given by a set of four differentiable, increasing and strictly quasi-concave functions,  $F^{++}(x)$ ,  $F^{+-}(x)$ ,  $F^{-+}(x)$ , and  $F^{--}(x)$ . Denoting their respective gradients:

$$f^{++}(x) = (f_1^+(x), f_2^+(x))$$

$$f^{-+}(x) = (f_1^-(x), f_2^+(x))$$

$$f^{+-}(x) = (f_1^+(x), f_2^-(x))$$

$$f^{--}(x) = (f_1^-(x), f_2^-(x))$$

it is assumed that, for each  $x \geq 0$ ,

$$f_1^-(x) > f_1^+(x) \tag{2.1}$$

$$f_2^-(x) > f_2^+(x) \tag{2.2}.$$

If the effect on anticipated utility of any move  $dx$  from any initial situation  $x \geq 0$  is such that:

$$dU = \begin{cases} f^{++}(x)dx & \text{if } dx_1, dx_2 \geq 0 \\ f^{-+}(x)dx & \text{if } dx_1 \geq 0, dx_2 \leq 0 \\ f^{+-}(x)dx & \text{if } dx_1 \leq 0, dx_2 \geq 0 \\ f^{--}(x)dx & \text{if } dx_1, dx_2 \leq 0 \end{cases}$$

this effect will be the same as the one which would be expected from conditions [1.1]-[1.2], that is, “pleasure of gaining is not equal to the evil of losing”.

It must be noted that although conditions [2.1]-[2.2] do not concern directly the marginal rates of substitution, but magnitudes usually interpreted as marginal utilities, they do

not imply anything like a cardinal utility function: it is obvious that the same conditions are satisfied if  $F^{++}(x)$ ,  $F^{+-}(x)$ ,  $F^{-+}(x)$ , and  $F^{--}(x)$  are submitted to any monotone increasing transformation.

Let us now take explicitly in consideration the *anticipated* nature of the move  $dx$ : anticipating  $dx$  means that at date  $t$ , we anticipate that at  $t + dt$  ( $dt > 0$ ), our situation will move to  $x(t) + x'(t)dt$  (where  $x'(t)dt = (\partial x / \partial t)dt = dx$ ). The anticipated variation of utility  $dU$ , at any initial situation  $x$ , generated by an anticipated move  $dx$ , can now be defined as:

$$dU = \varphi(x(t), x'(t))x'(t)dt \quad [2.3]$$

where  $\varphi(x(t), x'(t))$  is a function of sensitivity to gains and losses, such that:

$$\varphi(x(t), x'(t)) = \begin{cases} f^{++}(x) & \text{if } x'_1, x'_2 \geq 0 \\ f^{+-}(x) & \text{if } x'_1 \geq 0, x'_2 \leq 0 \\ f^{-+}(x) & \text{if } x'_1 \leq 0, x'_2 \geq 0 \\ f^{--}(x) & \text{if } x'_1, x'_2 \leq 0 \end{cases}$$

Sensitivity to gains and losses, described by equation [2.3], suggests that the variations of anticipated utility between an initial and a final date – say,  $t_0$  and  $t_f$  –, are obtained by integrating  $dU$  along a trajectory  $\{x(t)\}$ . Anticipated utility then appears as a function, not of a final situation  $x(t_f)$ , but of a trajectory  $\{x(t)\}$ . This means that an individual's ability to anticipate the consequences, on his utility, of a move  $dx = x'dt$  goes along with his ability to anticipate the consequences of a new move, once  $dx$  is achieved. However, note that anticipated utility should not be understood as effective utility enjoyed by an agent, but just as the anticipation, at  $t_0$ , of utility at  $t_f$ , which depends on the evolution of sensitivity to gains and losses along a trajectory. For convenience, we will admit that the anticipated utility of remaining in  $x(t_0)$  at date  $t_0$  is 0. Henceforth, anticipated utility of a trajectory  $\{x(t)\}$  is given by:

$$U(\{x(t)\}) = \int_{t_0}^{t_f} [\varphi(x(t), x'(t))x'(t)]dt \quad [2.4].$$

It must be stressed that  $\varphi(x(t), x'(t))$  is piecewise continuous. More precisely, discontinuity occurs at each date when the sign of at least one component of the right-hand derivative of  $x(t)$  – hence, of  $dx$  – changes<sup>10</sup>.

It is clear that if time is important here, it is only to the extent that it provides an index of the order of the situations anticipated along  $\{x(t)\}$ : it entails, in itself, neither utility nor

---

<sup>10</sup>The anticipated utility of a trajectory  $\{x(t)\}$  appears as a piecewise continuous function. For example, let  $x'_1, x'_2 > 0$  for  $t_0 \leq t < \theta$ , and  $x'_1 < 0, x'_2 > 0$  for  $\theta \leq t \leq t_f$ . Then,

$$\begin{aligned} U(\{x(t)\}) &= \int_{t_0}^{\theta} [f_1^+ x'_1 + f_2^+ x'_2]dt + \int_{\theta}^{t_f} [f_1^- x'_1 + f_2^+ x'_2]dt \\ &= (F^{++}(x(\theta)) - F^{++}(x(t_0))) + (F^{-+}(x(t_f)) - F^{-+}(x(\theta))). \end{aligned}$$

desutility. What matters is then the shape of the oriented trajectory  $\{x(t)\}$  in the space of quantities  $x_1$  and  $x_2$ ; but the time length between any points of the trajectory has no incidence on anticipated utility. This seems to be a rather strong assumption: in fact, it is not. Time between  $t_0$  and  $t_f$  is only time of choice, of possible individual deliberation. It is logically distinct from time involved in the definition of the quantities  $x_1$  and  $x_2$  – which might be of different dates –, or from time of the effective moves, which would eventually bring the individual from  $x(t_0)$  to  $x(t_f)$ . By analogy, if we were in a walrasian framework, a similar conception of time would have prevailed to give an account of the *tâtonnement* process, but of course neither to define goods, nor to describe actual exchange.

From the point of view of the trajectory, now,  $\{x(t)\}$  denotes the mental process which will make it clear which utility the agent might anticipate after performing in his mind each step of the trajectory. It seems possible to argue successfully that the trajectory should be linear – and, moreover, that a discrete-time presentation might be more appropriate. Indeed, if I imagine a transaction which would lead me from an initial to a final situation by giving up good 1 in exchange of good 2, the intermediate situations are meaningless, and the utility that I anticipate in  $t_f$  is  $F^{-+}(x(t_f)) - F^{-+}(x(t_0))$ ; it does not make sense for me to take into account such intermediary situation in which, for example, my endowment in good 1 would have increased, at the expense of my endowment in good 2. Quite different, of course, would be the case where an intermediary situation, for any reason, becomes significant. For instance, I could imagine exchanging good 1 for good 2, and then wonder how I would appreciate my previous initial situation, or some other situation which, from  $x(t_0)$ , was equivalent to the one I would have reached. But in such cases, one should observe that  $\{x(t)\}$  would present itself as a concatenation of linear trajectories.

Nonetheless, in spite of these arguments which induce to restrict meaningful  $\{x(t)\}$ 's to linear – or concatenations of linear – trajectories, we will consider, in the following, that  $\{x(t)\}$  might be any continuous trajectory.

## 2.2. Optimal trajectories and indifference maps

Given  $x(t_0) = x^0$ , the anticipated utility function hence assigns an utility index  $U(\{x(t)\})$  to any possible trajectory completed at any final situation  $x(t_f) \geq 0$ . It is then evident that this utility index does not depend exclusively on the final situation but, more generally, on the entire trajectory. This raises the question of the characterisation of optimal trajectories  $\{\hat{x}(t)\}$ , such that anticipating a move from  $x(t_0) = x^0$  to  $x(t_f) = x^f$  entails a maximum of anticipated utility. That is:

$$U(\{\hat{x}(t)\}) \geq U(\{x(t)\}) \quad [2.5]$$

where  $\hat{x}(t_0) = x(t_0)$  and  $\hat{x}(t_f) = x(t_f)$ .

Denote  $\{\tilde{x}(t)\}$  a monotone trajectory along which no strict sign reversal occurs concerning the first derivatives  $\tilde{x}'_1$  and  $\tilde{x}'_2$ . It is easy to show that monotone trajectories are equivalent to optimal trajectories (see annexe 1). As a result, linear trajectories or concatenations of linear trajectories in which no strict reversal occurs, are also equivalent to monopone trajectories.

On the basis of optimal trajectories, it becomes possible to represent the map of anticipated indifference between the final situations of optimal trajectories sharing the same starting point,  $x^0$ . Like in figure 1 below, each indifference curve is built with appropriate portions of the contours of  $F^{++}(x)$ ,  $F^{+-}(x)$ ,  $F^{-+}(x)$  and  $F^{--}(x)$ . For instance, let  $U(\{\hat{x}^1(t)\})$  denote the anticipated utility of an optimal trajectory leading from  $\hat{x}^1(t_0) = x^0$  to  $\hat{x}^1(t_f) = x^1 > x^0$ , and  $e_1, e_2$  the respective quantities of goods 1 and 2 for which  $F^{++}(e_1, x_2^0) = F^{++}(x_1^0, e_2) = F^{++}(x^1)$ . Then, the locus of the final situations of optimal trajectories granting the same anticipated utility as  $\{\hat{x}^1(t)\}$  is such that:

- if  $x_1 \leq x_1^0, x_2 \geq x_2^0$ ,  
 $x: F^{+-}(x) = F^{+-}(x_1^0, e_2)$
- if  $x_1 \geq x_1^0, x_2 \geq x_2^0$   
 $x: F^{++}(x) = F^{++}(x^1)$
- if  $x_1 \leq x_1^0, x_2 \leq x_2^0$   
 $x: F^{-+}(x) = F^{-+}(e_1, x_2^0)$ .

The resulting map of indifference, of course looks like the traditional one. However, four major differences should be noted:

1. The utility dealt with here, is anticipated utility, determined by a trajectory – not by an allocation.
2. The situations  $x$  represent the final points of optimal trajectories.
3. Although the indifference curves are convex to the origin, they show discontinuities in the marginal rate of substitution at each point of coordinates  $(x_1, x_2^0)$  or  $(x_1^0, x_2)$ .
4. The map of indifference is relative to the initial situation  $x^0$ .

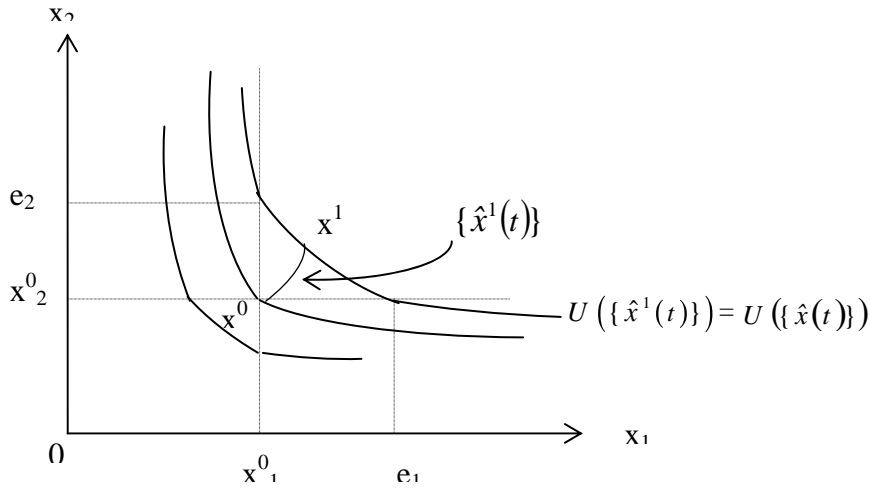


Figure 1: Anticipated indifference map relative to  $x^0$  between the final situations of optimal trajectories

This last remark has outstanding consequences. If, for any reason, one anticipates a move from  $x^0$  to  $x^1$ , and then from  $x^1$  to  $x^2$ , this will lead to consider  $x^1$  as a new initial situation, hence modifying the indifference map. Obviously, as the indifference maps relative to  $x^0$  and to  $x^1$  are constructed on the basis of the contours of the same functions, they share common parts. The indifference map is hence the same for all situations  $x$  such that the signs of  $x_i - x_i^0$  and  $x_i - x_i^1$  ( $i = 1, 2$ ) are the same (see figure 2). Otherwise, the agent might be faced with conflicting preference orders.

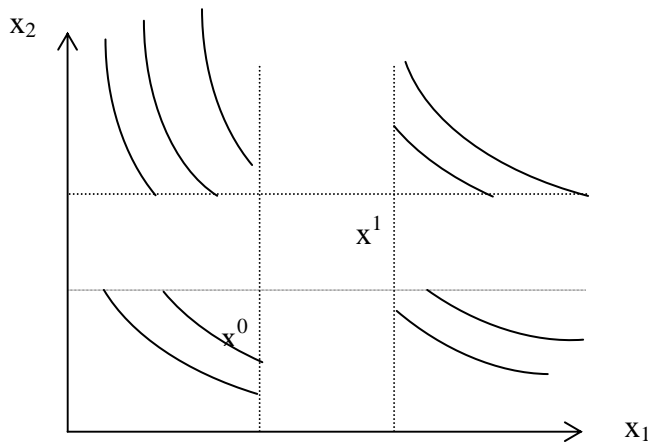


Figure 2: Common parts of the indifference maps relative to  $x^0$  and to  $x^1$

### 3. Conflicting preference orders

The coexistence of different preference orders is not as paradoxical as it seems. Acknowledging that individual's choices are governed by anticipated utility diminishes the

importance of final situations: the latter are only the last points of trajectories involving gains and losses, which determine completely anticipated utility. Nonetheless, even an individual guided by the Benthamite principle of asymmetric sensitivity to gains and losses might be led to consider with some perplexity the potential effects of a behaviour which would induce him to prefer, in one case  $x^A$  to  $x^B$ , and in an other case  $x^B$  to  $x^A$ . This justifies the importance granted to possible modifications of preference orders.

### 3.1. Preference reversal

#### 3.1.1. Preference reversal as a consequence of asymmetric sensitivity to gains and losses

The phenomena of preference reversal and, among them, of preference reversal between initial situations, as particular cases of preference order modifications, deserve special attention. Preference reversal between initial situations simply consists in the fact that the order of preferences between two potential initial situations, say  $x^0$  and  $x^1$ , depends upon the agent's initial situation, again  $x^0$  and  $x^1$ . It can be understood as a possible result of a comparison of the variations of anticipated utility associated with two couples of optimal trajectories<sup>11</sup>.

The first couple is made up with a *statu quo* trajectory, denoted  $\{x^{00}(t)\}$ , in which the agent anticipates remaining in  $x^0$  from  $t_0$  to  $t_1$ , and a trajectory  $\{x^{01}(t)\}$ , which leads him from  $x^{01}(t_0) = x^0$  to  $x^{01}(t_1) = x^1 \neq x^0$ . The second couple is formed in the same way, but the initial situation is now  $x^1$ . Henceforth,  $\{x^{11}(t)\}$  is the *statu quo* trajectory in which the agent stays in  $x^1$ , and  $\{x^{10}(t)\}$  is the trajectory which leads him to  $x^{10}(t_1) = x^0$ . Let  $\Delta_0$  and  $\Delta_1$  be the respective difference of anticipated utility between the non-*statu quo* and the *statu quo* trajectories. The *effective* initial situation of the agent could be any  $x^\theta$ , at date  $t_\theta \leq t_0$ .  $x^0$  and  $x^1$  thus appear as significant intermediary situations of more complex trajectories, made from the concatenation of optimal trajectories.  $\Delta_0$  (resp.  $\Delta_1$ ) then answers to the following question: imagine that the agent is in  $x^\theta$ , and that he anticipates an optimal move to  $x^0$  (resp.  $x^1$ ); how would he appreciate moving further to  $x^1$  (resp.  $x^0$ )? Clearly, preference reversal occurs when the signs of  $\Delta_0$  and  $\Delta_1$  are identical. Annexe 2 shows that such is the case when  $\Delta_0$  and  $\Delta_1$  are both negative (*i.e.* refer to possible deteriorations of the agent's situation, in terms of anticipated utility), under some restrictive conditions. More precisely, it shows that the condition is that  $x^1$  should be situated between the curves  $F^{+-}(x) = F^{+-}(x^0)$  and  $F^{-+}(x) = F^{-+}(x^0)$  (see figure 3). If such is the case, the agent will always prefer the *statu quo*: when he is in  $x^0$ , he

---

<sup>11</sup> As the context makes it clear, the “ $\wedge$ ” will be omitted hereafter to denote optimal trajectories.

anticipates a greater utility from remaining in  $x^0$  than from moving to  $x^1$ ; but when he is in  $x^1$ , he also infers a greater utility from remaining in  $x^1$  than from moving to  $x^0$ .

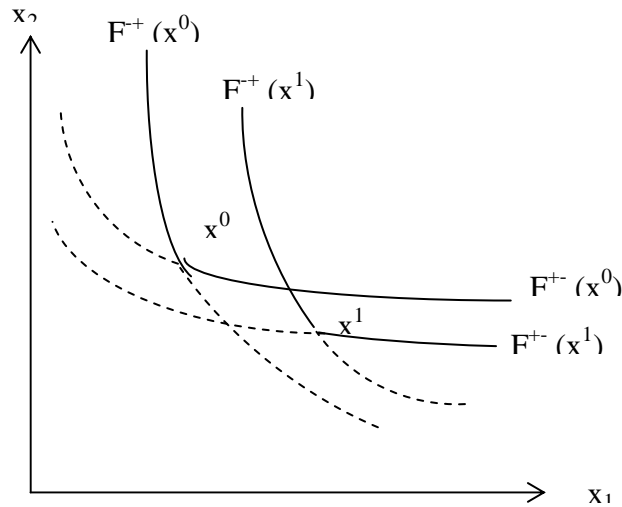


Figure 3: Preference reversal between initial situations  $x^0$  and  $x^1$

This conclusion can easily be extended to the general case of preference reversal. Annexe 3 gives the formal conditions for such a generalisation: it shows that under some restrictive conditions, given two possible initial situations  $x^0$  and  $x^1$  and two possible final situations  $x^A$  and  $x^B$ , the agent anticipates a greater utility from a move to  $x^A$  when he is, for instance, in  $x^0$ , and a greater utility from a move to  $x^B$  when he is in  $x^1$ .

### 3.1.2. To what extent are “preference reversals” paradoxical ?

The above interpretation of preference reversals should be distinguished from the traditional one, in the line of the pioneering paper of S. Liechtenstein and P. Slovic (1971). Usually, Liechtenstein’s and Slovic’s paradoxical result – the highest reservation price is set by a decision maker to the lottery which he does not prefer – was analysed in the context of a theory of choice in uncertainty, so that it could be linked to other types of non-expected utility behaviour, like the well-known Allais’ paradox. A primary reaction to the evidence of preference reversals provided by experimental data, was to spare usual choice theory by discarding the mechanism used to find the decision maker’s selling price of lotteries<sup>12</sup>. Meanwhile, the debate focused on a more critical area: at issue, was the attempt to explain Liechtenstein’s and Slovic’s result in terms of the violation of some fundamental axiom of the theory of choice, either the axiom of transitivity (Loomes and Sudgen 1982), or the independence axiom (Holt 1986) – this latter being directly linked to the linearity of the

<sup>12</sup> The first systematic analysis of Liechtenstein’s and Slovic’s paper, from an economic point of view, seems to have been D. Grether and C. Plott (1979), who tried to show that the authors’ paradoxical results were a consequence of the method in use, but that the main hypotheses of the theory of choice were not seriously compromised.



probabilities hypothesis in expected utility theory. Other contributions, like Quiggin (1982), modified the independence axiom and introduced a function of deformation of probabilities. The question is then to know whether violation of independence or transitivity is involved in preference reversals from a Benthamite point of view.

It might be argued that Bentham's integration of probability as a dimension of pleasure or pain in chapter IV of the *Introduction to the Principles of Morals and Legislation* opens the path to a new interpretation of choices in uncertainty. A remark from K. Arrow to M. Machina in a conversation (Machina 1987: 130, n. 10), according to which "individuals display more risk aversion in the event of an opportunity loss, and less risk aversion in the event of an opportunity gain" suggests what a reformulation of Bentham's position would look like: Arrow's interpretation might evidently be understood as an expression, in a risk context, of the Benthamite principle that the "pleasure of gaining is not equal to the evil of losing". This would be an alternative to, for example, Quiggin's decumulative probabilities, since it concerns directly not the perception of probabilities, but – like in Friedman and Savage (1948)'s article – the perception of utility. However, this possibly fruitful extension of the Benthamite asymmetry between gains and losses will be deliberately neglected in the remainder of this paper. As a consequence, independence is clearly offside.

Does this mean that transitivity is transgressed ? Of course, not: indeed, it is not  $x^A$  which is preferred to  $x^B$  in one case, and  $x^B$  which is preferred to  $x^A$  in the other case, but two pairs of different trajectories leading to the same pair of final situations, but associated to different anticipated utility. Nothing, in this, seems to challenge the idea that preferences *over trajectories* are transitive. If "preference reversal" nonetheless appears paradoxical, it is only because we neglect the fact that  $x^A$  and  $x^B$  are each one introduced in the reasoning as the final situations of different trajectories, which are the true arguments of the function of anticipated utility.

Of course, this is puzzling, since the difference in anticipated utility rests neither on any difference in the final outcome of the process, nor on any cost involved in effective transactions, but on the individual's mental process of exploring alternative decisions. Nonetheless, in contrast to the modern preference reversal approach, the Benthamite preference reversal cannot be considered as a threat against the independence or the transitivity axiom.

### **3.2. An instance of what Bentham said : wage, effort, premium, penalty and the labour contract**

Although the preceding results concerning preference reversal appear as a logical extension of the Benthamite principle, according to which the "pleasure of gaining is not

equal to the evil of losing”, we might wonder whether Bentham was or not clearly aware of these analytical consequences of his writings. However, the thorough investigation concerning the labour contract, to which are devoted numerous pages of the *Rationale of Reward*, witnesses on the one hand, his awareness of the possible implications of his so often recalled conceptions and, on the other hand, that access to the subtleties of differential and integral calculus was – at Bentham’s time – superfluous to grasp some major consequences of a psychological assumption like asymmetric sensitivity to gains and losses.

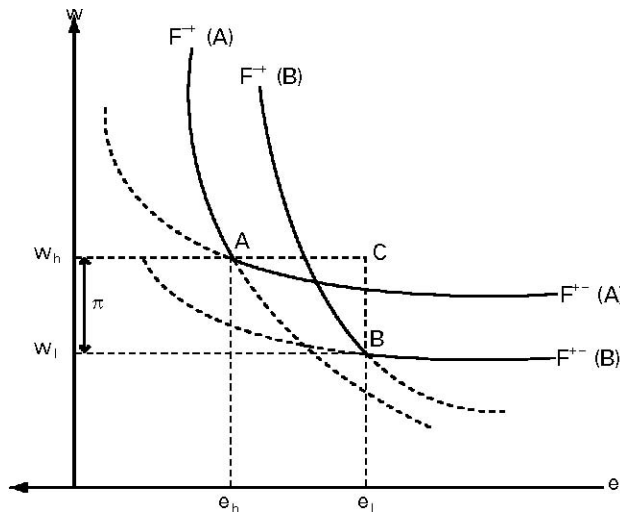
Like the *Manual of Political Economy*, the *Rationale of Reward* is an excerpt of a more ambitious work, published in French by Etienne Dumont under the title *Théorie des Peines et des Récompenses* (1811). In the “Preliminary Observations” to this last work, Bentham denounced the illusion of a formal symmetry between pain and pleasure, arguing at length that, from the point of view of the legislator, the threat of a penalty was much more efficient, as an individual incentive, than the promise of a reward. This primary principle was then applied to the analysis of the remuneration attached to public offices.

In fact, Bentham’s main point was to stress that, in most cases, wages did not reward anything: they did not ensure assiduity, nor application, and the performance of the minimal work normally required solely depended on the goodwill of their beneficiary. The problem of control of economic activities was then raised, at a time when it was far from being a worthwhile matter to the economists who were Bentham’s contemporaries. Meanwhile, he did not really suspect that the difficulty of control was an effect of the private character of the information held by individuals. More precisely, Bentham surely acknowledged that information was costly – this justified the existence of inspectors – but he did not imagine that it might be really private: one just has to go and see if the task had been performed. The difficulty of control then stems from the organisation of labour: if wages are a reward neither for the employee, nor for the inspector, the work has no chance to be correctly achieved. Bentham’s purpose was hence to provide appropriate incentives in order that work – including control – is performed efficiently. These incentives – in the particular meaning granted to the word here, since there is no hidden information to manage – chiefly consist in the form of the labour contract.

Bentham’s recommendation can easily be summarised: in order that the demanded effort be accomplished, give proper wages to the labourer, but submit them to the condition that he will be inflicted a penalty if the task is not correctly performed. Curiously, he did not favour the symmetric contract – which would stipulate low wages, possibly increased by a premium if the task is correctly performed – since it seems unable to induce the worker to

perform his job with all the necessary skill<sup>13</sup>. Of course, the foundations of such a position seem rather mysterious to a modern economist, because the two contracts open the path to the same set of choice: a) high wages and high effort, and b) low wages and low effort. And if only final situations are taken into account for determining choices, there is no reason for a) to be chosen in one case, and b) in the other.

But Bentham's vindication of the first type of contract becomes understandable if asymmetric sensitivity to pain and pleasure is considered, and if utility is perceived as anticipated utility, that is as the outcome of a full trajectory leading either to high wages and effort, or to low wages and effort. To precise the argument, assume, like in figure 4, that these final situations are denoted  $A = (e_h, w_h)$ , and  $B = (e_l, w_l)$  –  $e$  and  $w$  being respectively the effort performed by the worker and his wage rate, and the subscripts  $h$  and  $l$  standing for “high” and “low”. Assume also that  $F^{-+}(B) > F^{-+}(A)$  and  $F^{+-}(A) > F^{+-}(B)$ . This is clearly a condition for preference reversal on initial situations, as described in § 3.1. Imagine then that at date  $t = t_0$ , the worker is in the situation denounced by Bentham, that is, that he earns high wages and, since his work is not submitted to any control, that he performs a low effort. On figure 5, point  $C = (e_l, w_h)$  corresponds to this situation. Denoting  $\{x^{CC}(t)\}$ ,  $\{x^{CA}(t)\}$ , and  $\{x^{CB}(t)\}$  the optimal trajectories which respectively lead the worker from  $C$  to  $C$ ,  $A$ , and  $B$ , it is obvious that  $U(\{x^{CC}(t)\}) > U(\{x^{CA}(t)\})$  and  $U(\{x^{CC}(t)\}) > U(\{x^{CB}(t)\})$ , but it is not absolutely clear which is the greater, from  $U(\{x^{CA}(t)\})$  and  $U(\{x^{CB}(t)\})$ <sup>14</sup>.



( $e$  : inverted scale)

Figure 4: The Benthamite labour contract

<sup>13</sup> See Bentham 1782-7: 236: “It is by fear only and not by hope, that [a worker] is impelled to the discharge of his duty - by the fear of receiving less than he would otherwise receive, not by the hope of receiving more”.

<sup>14</sup> However, Bentham seems to advocate the idea that usually, the decrease in wages generates a higher pain than the increase in effort, so that in most cases,  $U(\{x^{CA}(t)\}) > U(\{x^{CB}(t)\})$ .

As long as no institutional constraint has the effect of withdrawing  $C$  from the set of choice of the worker, there is no inducement for his giving up  $C$  in favour of  $A$  or  $B$ . But suppose that such a reform has been instituted, through the conditions of the labour contract. Since preference reversal occurs between  $A$  and  $B$ , the worker will always anticipate a greater utility from an optimal move to  $A$ , than from the same move, followed by a second optimal move to  $B$ . Symmetrically, he will prefer moving to  $B$ , than moving to  $A$  after having moved to  $B$ . Hence, when Bentham advocates a labour contract of the type “high wages and conditional penalty”, he proposes the institutional device which places the worker in initial situation  $A$ . And, if the penalty is  $\pi = w_h - w_l$ , it is clear that the agent is not willing to move to  $B$ , and prefers *statu quo*:

$$U(\{x^{AA}(t)\}) > U(\{x^{AB}(t)\}).$$

Conversely, had the labour contract been of type “low wages and conditional premium”, this would have placed the agent in  $B$ , thus compromising his willingness to increase his income by means of his effort – that is, to move to  $A$ :

$$U(\{x^{BB}(t)\}) > U(\{x^{BA}(t)\}).$$

Whereas, at first sight, Bentham’s analysis of the labour contract seems rather obscure, prisoner of technical and legal considerations, tedious even for the most lenient reader, it nevertheless shows to be a skilful proposal, which takes advantage of the worker’s preferences, in order to induce him to accept both to perform a high effort, and to be rewarded for this. Moreover, Bentham’s clear understanding of such analytical consequences of asymmetric sensitivity to gains and losses as preference reversal, becomes apparent behind his confidence in one type of contract at the detriment of the other.

### **3.3. An instance of what Bentham didn’t say: demand, equilibrium, and the existence of anticipated non-exchange situations**

Up to our knowledge, there is no explicit reference, in Bentham’s writings, to the link between anticipated utility and the identification of some kind of prefiguration of equilibrium. In the same way, no mention can be found of the relation with a function of demand. Nonetheless, equilibrium and demand appear – at least, to a modern economist – as natural extensions of utility theory. It is therefore not surprising that the same analytical device which made clear Bentham’s intuition concerning the specific case of the labour contract, might also give birth to more general statements.

The question of individual equilibrium can be dealt with in the framework of a simple representation of consumer behaviour, submitted to a linear budget constraint. Assume that,

from  $x^0 = x(t_0)$ , the agent anticipates that at  $t = t_1$ , prices will be  $p^1 = p(t_1)$ . Maximising anticipated utility hence comes down to finding an optimal trajectory  $\{x^*(t)\}$  solution of

$$\left| \begin{array}{l} \max U(\{x(t)\}) = \int_{t_0}^{t_1} [\varphi(x(t), x'(t))x'(t)] dt \\ \text{subject to: } p(t_1)x^0 - p(t_1)x(t_1) = 0 \end{array} \right. \quad [3.1].$$

Considering that the final situation  $x^*(t_1)$  of  $\{x^*(t)\}$  is an equilibrium, it can be conveniently obtained by resolving static maximisation problems (see figure 5):

- (a) If  $p_1(t_1)/p_2(t_1) > f_1^-(x^0)/f_2^+(x^0)$  – that is, if the price ratio is greater than the left-hand marginal rate of substitution – the agent is induced to supply good 1 and to demand good 2, so that  $x^*(t_1)$  is solution of

$$\left| \begin{array}{l} \max F^{+-}(x) \\ \text{subject to: } p(t_1)x^0 - p(t_1)x = 0 \end{array} \right. \quad [3.2].$$

- (b) If  $f_1^-(x^0)/f_2^+(x^0) \geq p_1(t_1)/p_2(t_1) \geq f_1^+(x^0)/f_2^-(x^0)$ , it is clear that the agent is submitted to no incentive to move from  $x^0$ , which appears as an equilibrium situation:

$$x^*(t_1) = x^0.$$

- (c) If  $f_1^+(x^0)/f_2^-(x^0) > p_1(t_1)/p_2(t_1)$ , the agent would improve his situation by exchanging good 2 for good 1, so that  $x^*(t_1)$  is now solution of

$$\left| \begin{array}{l} \max F^{+-}(x) \\ \text{subject to: } p(t_1)x^0 - p(t_1)x = 0 \end{array} \right. \quad [3.3].$$

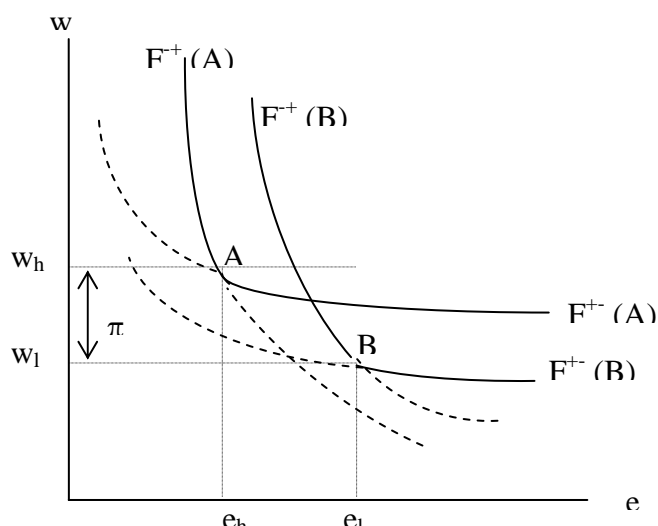


Figure 5: Individual equilibrium

According to the relative positions of the anticipated price ratio and of the marginal rates of substitution, the net demand for good 1 or 2 will be (figure 6):

- elastic to prices and derived from [3.2] in case (a);
- inelastic and equal to zero when prices belong to the close interval of case (b);
- elastic to prices and derived from [3.3] in case (c).

The existence of an interval of prices within which the demand function is inelastic and the demand is zero, has a significant consequence from the moment it is supposed that  $x^*(t_1)$  is only an intermediary situation. In  $x^*(t_1)$ , the map of indifference has changed, so that, for instance, if  $p_1(t_1)/p_2(t_1) = f_1^+(x^*(t_1))/f_2^-(x^*(t_1)) > f_1^+(x^0)/f_2^-(x^0)$ , a small enough increase in the price ratio at  $t = t_2$  has no effect on demand: this latter remains  $x^*(t_1) - x^0$ , even if the same increase, at  $t = t_1$  would have provided no incentive for the agent to anticipate a move from  $x^0$ . This remark helps qualifying the interpretation of the demand function, such as defined above. Its specificity is not only to be partly price-inelastic. The link that it shows between net demand and anticipated price is subordinated to the fact that the demanded quantities are associated with the final situations of *monotonous* trajectories from  $x^0$ . Otherwise – typically when  $p$  follows an arbitrary path  $\{p(t)\}$  – the agent might demand different quantities of goods 1 and 2 for the same relative prices at  $t_1$  and  $t_2$ , according to the trajectory he anticipates to perform between  $t_1$  and  $t_2$ .

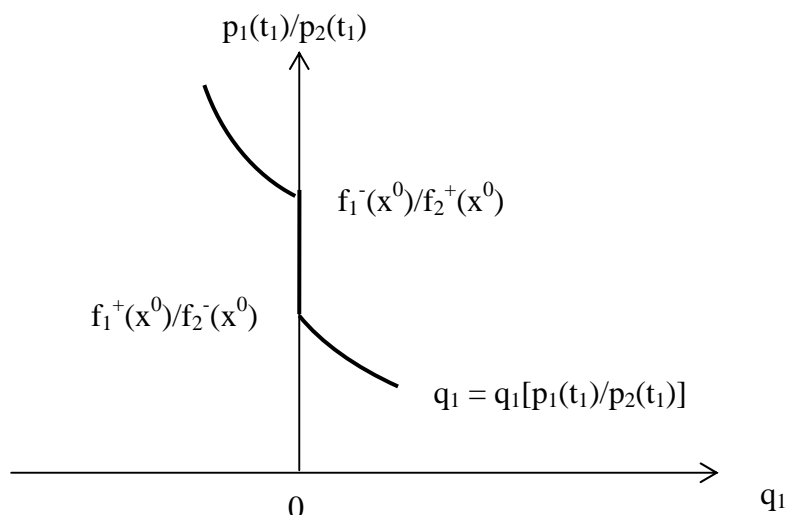


Figure 6: Net demand function for good 1

The three kinds of solutions to the dynamic maximisation program [3.1] play a determining part in the characterisation of general equilibrium. It can be shown (see annexe 4) that the usual contract curve moves into a compact contract set. From any initial situation

exterior to the contract set, some relative prices allow an anticipated general equilibrium exchange on its border, so that the agents positive demands and supplies are all satisfied. On the contrary, within the contract set, general equilibrium is obtained only by the relative prices which ensure that each agent has zero net demands and supplies. This obviously means that each initial situation in the contract set is also the general equilibrium relative to itself.

On first view, the above construction only shows poor compatibility to the strict general equilibrium approach. The reason is that preferences depend on an anticipated trajectory, whereas it is usually admitted that they are given in general equilibrium. However, the opposition is not as clear-cut as it looks to be.

In the traditional case, it is by seek of convenience that we say that preferences are “exogenous”. In fact, general equilibrium theory only grasps, at  $t_0$ , a certain state of preferences – even intertemporal –, technology and endowment, and identifies the price vector which makes them mutually consistent. Nothing, in the theory, excludes that at any  $t > t_0$ , preferences might have changed, involving a revision of equilibrium. In other words, preferences might be considered exogenous in a general equilibrium context, only to the extent that no interest is borne to what happens i) after  $t_0$  ; and ii) before  $t_0$  – except, of course, to the *tâtonnement* process.

Now, it is clear that whereas the anticipated utility assumption does not require any attention to be paid to effective future events – that is, to what happens after the identification of equilibrium prices and trajectories – it focuses on the mental process which precedes the final choice and, moreover, any demand or supply expressed along a *tâtonnement* process. From this point of view, it brings to the forefront a “time of deliberation” which might rightly be neglected in traditional theory, since this latter assumed that no asymmetric sensitivity to gains and losses occurs. Therefore, the anticipated utility assumption is less an alternative to a general equilibrium approach, than an attempt to investigate what allows us to acknowledge that, from a certain moment, we are entitled to admit that preferences are given.

#### **4. From the time of assessment to the time of deliberation**

The mental process to which attention has been paid until now unfolds from the identification of the potential pleasures and pains associated to the set of achievable trajectories, to the effective decision which should follow it. Bentham’s “*felicific calculus*” then appears as an attempt to give an account of this process.

It must be underlined that the *felicific calculus* covers two distinctive operations: i) the assessment of anticipated utility – of pleasure and pain – associated to each trajectory, and ii) the deliberation which leads from a simple comparative assessment of utility towards a

decision involving a choice. No doubt that in Bentham's mind, the time of assessment and the time of deliberation were melted. However, his meticulous discussion of both operations helps distinguishing them.

#### **4.1. The time of assessment: what does it mean, being a bad calculator ?**

The time of assessment is concerned in order to explain what would be, from a Benthamite point of view, a "bad calculator". This latter is characterised by his poor ability to assess correctly the consequences of his acts in terms of utility – that is, to use efficiently all available information. Such would be, for instance, a criminal that cannot be deterred in his enterprises, even by the most severe perspective of punishment. In terms of the formalization of the previous sections, this means that the individual considers as optimal an anticipated trajectory which *is not*. Now, this point deserves special attention. Indeed, this "is not" is rather puzzling, since it echoes the juncture between the positive and the normative aspects of the principle of utility. Of course, it does not imply that the agent, after anticipating a trajectory, anticipates an other trajectory, that would lead him to a new final situation which both improves his anticipated utility, and was previously reachable: in other words, it has nothing to do with the possible existence of conflicting preference orders. Nor is it the result of the awareness, after the trajectory is achieved, that this latter was sub-optimal: after all, even if his enterprise fails, the criminal might persist in his error – that is, in his bad assessment. The fact that an alleged optimal trajectory "is not" optimal simply means that the agent's calculation is not consistent with his tastes. However, since we implicitly admitted until now that the description of tastes is nothing else but the calculation of anticipated utility generated by trajectories, the break between tastes and calculation comes down to turning to an external reference, some kind of picture of the agent, with the same tastes, but with a perfect knowledge of these tastes.

It is clear that the normative aspect of Bentham's principle of utility rests on this duality. Whereas individual agents act according to what they believe to be their optimal trajectories, the Benthamite legislator aims at taking into account what optimal trajectories would be in case of perfect knowledge of the tastes, and compares them to the actual behaviour of the individuals. This helps explain the reasons why Benthamite economics is, from the beginning, public economics: public intervention is, of course, required in order to provide adequate information, to make up for external effects and coordination failures (see Sigot 1993); but it is also required to ensure the internal consistency of each individual – that is, the consistency between his tastes and his resulting choices.

Several aspects of Bentham's legal reflection testify the permanency of his interest on the difficulties which come out during the time of assessment. His already mentioned involvement in an educational project (*supra*, § 1.2.) illustrates this point. Similarly, the first



“rule as to emoluments” (1782-7: 237-9), which advocates an “intimate connection between the duty and the interest of the person employed”, makes obvious that Bentham considered that reducing the temporal gap between associated pleasures and pains helps improve the workers’ ability to estimate the consequences of their actions in terms of utility. Numerous other examples – like his involvement in the debates concerning the “Poor Laws” – attest to Bentham’s effort to identify the source of this insufficient ability, and to provide proper solution. Evidently, this brings to the forefront the important question of the capacity of his recommendations to move each individual into a good calculator.

Self-enforcing mechanisms – illustrated by what Bentham called “self-executing laws” (*Ibid.*: 199-200) – play here a crucial part. Except in the case of direct constraint, of which the “Panopticon” is a typical instance, the utilitarianist educational project, in its broad sense, calls for each individual own interest, and therefore for his participation. For instance, if one knows that the ability to identify an optimal trajectory consistent with tastes is improved by learning, one is induced to learn how to identify this trajectory, even when this latter is not clearly anticipated. In other words, Bentham suggests that some transformation of an individual’s set of choice might give him the opportunity to apply *felicific calculus* as efficiently as possible.

#### **4.2. The time of deliberation: resolving the conflict between preference orders**

Let us then assume that these recommendations are implemented, and that they are a success. Imagine that the required information is available and that, owing to the wisdom of a Benthamite legislator, the discrepancy between tastes and calculation has disappeared from the surface of the Earth. Coming after assessment, deliberation would hence appear as the simple acknowledgement of a non-disputable optimal trajectory – that is, of the action for which the balance between pleasure and pain is the greatest. The resulting picture of an individual whose tastes are perfectly consistent with choices, and the well-known Veblenian caricature of the neo-classical economic agent, are now as like as two peas. It seems to be the end of the story.

It is not. What we described in § 2 and 3 assumes that the time of assessment is over, and proposes a rather elementary investigation into deliberation. The investigation is elementary because it supposes that, given an initial situation, the individual will follow an optimal trajectory, even if his preference orders are conflicting. Let us focus on this point. The Benthamite asymmetry between pleasure and pain rests on a special type of sensitivity – namely, that “the pleasure of gaining is not equal to the evil of losing”. The *statu quo* property of these conflicting preference orders (*supra* § 3.1.1. and 3.1.2.) implies that when an agent has achieved an optimal trajectory, no other final situation, previously reachable and preferred

from another initial situation, would ever seem better to him. Now, from an associationist point of view, it might be argued that, when you meet successively the same initial situation, the “copies of sensations” remain of the same nature, do not give birth to new ideas, so that no true deliberation is required to lead you straight from your preferences to your choice. There is, of course, such a position in Bentham’s works – especially when he deals with civil servants, stiffened by a rigid legislation.

However, as far as civil servants are no more concerned, the repetition of an identical initial situation becomes exceptional. The same associationist argument hence leads one to quite different conclusions: experimenting various initial situations now brings about different “copies of sensations”. In such a case, the individual knows what his optimal trajectory would be. But he is also aware that, starting from another initial situation, he would have preferred a situation that he now does not prefer. Moreover, it is even not necessary to assume *effective* changes in the initial situation: the ability to anticipate various trajectories comes with the ability to anticipate various initial situations. Henceforth, the assumption of correct assessment also means that the individual might be faced with conflicting preference orders. Resolving this conflict is therefore the stake of deliberation.

Again, the example of the Benthamite labour contract, examined above (§ 3.2), suggests some indications concerning what Bentham had in mind. On the one hand, it is possible to argue that the legislator who implements a contract associating high wages with a conditional penalty – instead of a contract stipulating low wages, possibly increased by a premium if the task is correctly performed – is chiefly concerned with the Benthamite objective of “abundance”, requiring a high level of production. But, on the other hand, the legislator knows that he does not really hurt any worker’s direct interest. Since the set of choices among final situations is the same for both types of contracts, the workers, like the legislator, are induced to neglect the effects of asymmetric sensitivity to pleasure and pain, in favour of utility associated to final situations. Formally, this means that utility associated to a final situation can no more be obtained by integrating along a trajectory, like in equation [2.4] of § 2<sup>15</sup>. The expected support of the workers in favour of the new labour contract shows that the legislator is not a dictator; but it also shows that there is no definite break between both conceptions of utility since, as a result from their deliberation, workers can persuade themselves that this contract is a solution to their conflicting preferences.

Of course, the outcome of this deliberation will not be the same for all individuals. Not only some of them remain sensitive to trajectories, but even in the case when this sensitivity is

---

<sup>15</sup> One of us argued that the difference between the utility of a situation and the utility of a trajectory might be rooted in Bentham’s dual conception of the *felicific calculus*. These two conceptions are respectively presented in chapters IV and V of the *Introduction to the Principles of Morals and of Legislation*. See Sigot 1995.

given up, the relative weights of goods in a final situation might more or less reflect the fear of a loss or the hope of a gain <sup>16</sup>. This analysis explains the importance of the part played by different types of characters within the Benthamite project – such as, for instance, the “prudent and well-grounded projector” praised in the *Defence of Usury*. Far from being an accessory illustration, psychological investigation in Bentham’s writings satisfies a theoretical requirement: the principle of utility sometimes shows one single path to follow; but more often, as a consequence of what Bentham called an “axiom of mental pathology”, it suggests that several paths might be practicable. Choosing one of them is hence no more a question of rationality, but of character. At least one century ago, for this very reason, the time of deliberation which achieves Bentham’s *felicific calculus* seems to have fled from the competence of economists. But there is no evidence that it is out of reach.

---

<sup>16</sup> The “utility of a situation” might therefore be analysed as a special case in which the function  $\varphi(\cdot)$  of sensitivity to gains and losses, as defined in § 2.1, does not depend any more on the signs of the elements of  $x$ . Since the individual becomes insensitive to gains and losses, it is obvious that  $f_i^+ x(t) \leq \varphi_i(x(t)) \leq f_i^- x(t)$ .

### Annexe 1: Equivalence between monotone and optimal trajectories.

Let  $\{\tilde{x}(t)\}$  and  $\{x(t)\}$  be respectively a monotone and a strictly non-monotone trajectory, both beginning in  $\tilde{x}(t_0) = x(t_0)$ , and finishing in  $\tilde{x}(t_f) = x(t_f)$ . Anticipated utilities of  $\{\tilde{x}(t)\}$  and  $\{x(t)\}$  can be written:

$$U(\{\tilde{x}(t)\}) = \int_{t_0}^{t_f} [\varphi(\tilde{x}(t), \tilde{x}'(t))\tilde{x}'(t)] dt \quad [\text{A11}]$$

and:

$$U(\{x(t)\}) = \int_{t_0}^{t_f} [\varphi(x(t), x'(t))x'(t)] dt \quad [\text{A12}].$$

As the signs of  $\tilde{x}'_1$  and  $\tilde{x}'_2$  do not change between  $t_0$  and  $t_f$ ,  $\varphi(\tilde{x}(t), \tilde{x}'(t))$  is continuous, as would be a function  $\varphi(x(t), \tilde{x}'(t))$ <sup>17</sup>.  $\{\tilde{x}(t)\}$  and  $\{x(t)\}$  having the same initial and final situations, the anticipated utility of the former might equivalently be expressed as:

$$U(\{\tilde{x}(t)\}) = \int_{t_0}^{t_f} [\varphi(x(t), \tilde{x}'(t))x'(t)] dt$$

[A13].

Consequently, the difference between anticipated utilities is:

$$U(\{x(t)\}) - U(\{\tilde{x}(t)\}) = \int_{t_0}^{t_f} [(\varphi(x(t), x'(t)) - \varphi(x(t), \tilde{x}'(t)))x'(t)] dt.$$

Now, as  $\{x(t)\}$  is assumed strictly non-monotone, there is necessarily at least one interval of time  $[t_1, t_2[$  within which the signs of at least one component of  $x'$  and  $\tilde{x}'$  – say,  $x'_i$  and  $\tilde{x}'_i$  – are different, so that  $\varphi_i(x(t), x'(t)) - \varphi_i(x(t), \tilde{x}'(t)) \neq 0$ . Given [2.1] - [2.2] as defined in §2.1 above,

- if  $\tilde{x}'_i > 0$ ,  $x'_i < 0$ , then:

$$(\varphi_i(x(t), x'(t)) - \varphi_i(x(t), \tilde{x}'(t)))x'_i(t) = (f_i^-(x(t)) - f_i^+(x(t)))x'_i(t) < 0$$

- if  $\tilde{x}'_i < 0$ ,  $x'_i > 0$ , then:

$$(\varphi_i(x(t), x'(t)) - \varphi_i(x(t), \tilde{x}'(t)))x'_i(t) = (f_i^+(x(t)) - f_i^-(x(t)))x'_i(t) < 0.$$

In both cases,  $(\varphi(x(t), x'(t)) - \varphi(x(t), \tilde{x}'(t)))x'(t)$  is therefore negative for  $t \in [t_1, t_2[$ , which implies that  $U(\{x(t)\}) - U(\{\tilde{x}(t)\})$  is also negative. As a result, all monotone trajectories from  $x(t_0)$  to  $x(t_f)$  are optimal, and all the strictly non-monotone trajectories are non-optimal. As a result, it should be noted that inequation [2.5] in §2.2 might then be strengthened:  $U(\{\hat{x}(t)\}) > U(\{x(t)\})$  when  $\{\hat{x}(t)\}$  is monotone and  $\{x(t)\}$  is strictly non-monotone.

---

<sup>17</sup> The function  $\varphi(x(t), \tilde{x}'(t))$  might be interpreted as a function of sensitivity to gains and losses, under this special assumption that, although the agent anticipates a trajectory  $\{x(t)\}$ , he keeps the same sensitivity to gains and losses as the one which prevailed all along  $\{\tilde{x}(t)\}$ .

## Annexe 2: The conditions for preference reversal

Let  $\Delta_0$  and  $\Delta_1$  be the respective difference of anticipated utility between the non-*statu quo* and the *statu quo* trajectories. That is,

$$\Delta_0 = U(\{x^{01}(t)\}) - U(\{x^{00}(t)\}) = \int_{t_0}^{t_1} [\varphi(x^{01}(t), x^{01'}(t))x^{01'}(t)] dt \quad [\text{A21}]$$

$$\Delta_1 = U(\{x^{10}(t)\}) - U(\{x^{11}(t)\}) = \int_{t_0}^{t_1} [\varphi(x^{10}(t), x^{10'}(t))x^{10'}(t)] dt \quad [\text{A22}]$$

Let us first examine the meaning and properties of the sum of  $\Delta_0$  and  $\Delta_1$ . Obviously, it represents the variation of anticipated utility generated by the successive performance of  $\{x^{01}(t)\}$  and  $\{x^{10}(t)\}$ , that is by a non-optimal trajectory  $\{x^{010}(t)\}$ , made up with an optimal move from  $x^0$  to  $x^1$ , followed by an other optimal move, going back to  $x^0$ . Given [2.5] (above, §2.2),  $\{x^{00}(t)\}$  being, itself, an optimal trajectory which gives birth to a zero variation of anticipated utility,  $\{x^{010}(t)\}$  is necessarily associated with a negative variation of anticipated utility. Thus,

$$\Delta_0 + \Delta_1 < 0 \quad [\text{A23}]$$

Since  $\{x^{01}(t)\}$  and  $\{x^{10}(t)\}$  are both optimal trajectories, [2.5] and [A23] imply that *any* trajectory, going from  $x^0$  to  $x^0$  through  $x^1$ , generates a decrease in anticipated utility, at most equal to  $\Delta_0 + \Delta_1$ . This conclusion is, of course, intuitive, since departing from  $x^0$  and coming back to it means that something has been gained and lost, involving more pain than pleasure.

Turning back to preference reversal, [A23] also implies that this latter only occurs when  $\Delta_0$  and  $\Delta_1$  are both negative. It follows that  $x^1$  can be neither greater nor smaller than  $x^0$ . We might then conclude that a necessary and sufficient condition for preference reversal is that  $x^1$  satisfies:

$$\left\{ \begin{array}{l} \Delta_0 = F^{+-}(x^1) - F^{+-}(x^0) < 0 \\ \text{and} \\ \Delta_1 = F^{-+}(x^0) - F^{-+}(x^1) < 0 \end{array} \right\} \text{ or } \left\{ \begin{array}{l} \Delta_0 = F^{-+}(x^1) - F^{-+}(x^0) < 0 \\ \text{and} \\ \Delta_1 = F^{+-}(x^0) - F^{+-}(x^1) < 0 \end{array} \right. \quad [\text{A24}]$$

### Annexe 3: The general case of preference reversal

Assume that from  $x^0$ , two optimal trajectories  $\{x^{0A}(t)\}$  and  $\{x^{0B}(t)\}$  lead respectively to  $x^A$  and  $x^B$ , so that

$$U(\{x^{0A}(t)\}) > U(\{x^{0B}(t)\}) \quad [\text{A31}].$$

If there be a situation  $x^1$ , from which two other optimal trajectories,  $\{x^{1A}(t)\}$  and  $\{x^{1B}(t)\}$  also lead to  $x^A$  and  $x^B$ , so that

$$U(\{x^{1B}(t)\}) > U(\{x^{1A}(t)\}) \quad [\text{A32}],$$

it is considered that the shift from the ancient to the new initial situation generates preference reversal between  $x^A$  and  $x^B$  (figure A1).

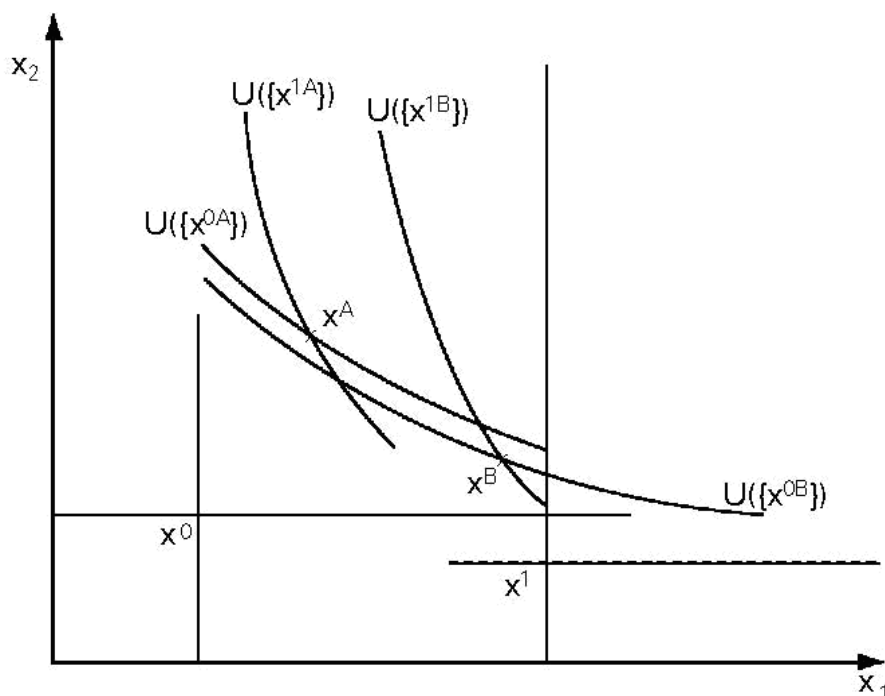


Figure A1: Preference reversal between  $x^A$  and  $x^B$  generated by a shift of the initial situation from  $x^0$  to  $x^1$

Suppose, now, a non-optimal trajectory  $\{x^{0AB}(t)\}$ , in which the agent anticipates an optimal move from  $x^0$  to  $x^A$ , and then an other optimal move from  $x^A$  to  $x^B$ . The anticipated utility of this trajectory is:

$$U(\{x^{0AB}(t)\}) = U(\{x^{0A}(t)\}) + U(\{x^{AB}(t)\}).$$

From [2.5], it is deduced that  $U(\{x^{0AB}(t)\}) < U(\{x^{0B}(t)\})$  and, if [A32] is true, that:

$$U(\{x^{AB}(t)\}) < 0.$$

In the same way, considering a non-optimal trajectory  $\{x^{1AB}(t)\}$ , we might conclude from [A32] that:

$$U(\{x^{BA}(t)\}) < 0.$$

Hence, if [A31] and [A32] are simultaneously verified, the locations of  $x^A$  and  $x^B$  necessarily imply:

$$\left\{ \begin{array}{l} F^{-+}(x^A) > F^{-+}(x^B) \\ \text{and} \\ F^{+-}(x^B) > F^{+-}(x^A) \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{l} F^{-+}(x^B) > F^{-+}(x^A) \\ \text{and} \\ F^{+-}(x^A) > F^{+-}(x^B) \end{array} \right. \quad [\text{A33}]$$

We might then conclude that, given [A31], if  $x^A$  and  $x^B$  relative positions satisfy conditions [A33] some  $x^1$  (for example,  $x^1 = x^B$ ) allowing [A32] – that is, preference reversal – could be found.

#### Annexe 4: The contract set and the characterisation of general equilibrium

Let us consider the simple case of a bilateral exchange between two agents,  $A$  and  $B$ , as represented in the box-diagram of figure A2.

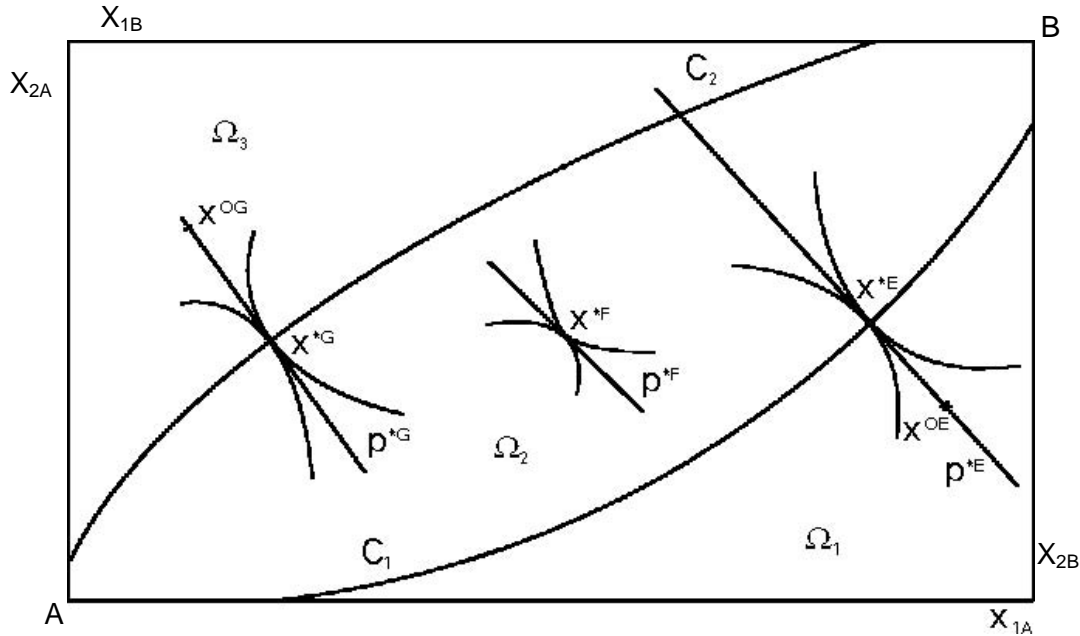


Figure A2: Contract set and general equilibrium

A first observation should be made. Assume that for prices  $p(t_1)$ ,  $x^0$  is not an equilibrium for both  $A$  and  $B$ , but there exists a final situation  $x^1$  which is an equilibrium. The inducement to move towards  $x^1$  is then an inducement to proceed to symmetric exchanges, so that when  $A$ 's maximisation program is [3.2],  $B$ 's is [3.3] (as defined above in §3.3), and reciprocally. Henceforth, we are faced with two contract curves  $C_1$  and  $C_2$  – instead of one, in the usual case – defined as the respective loci of the points of tangency of the contours of  $F_A^{++}(x)$  and  $F_B^{--}(x)$ :

$$C_1 : f_{A1}^-(x)/f_{A2}^+(x) = f_{B1}^+(x)/f_{B2}^-(x)$$

and of  $F_A^{+-}(x)$  and  $F_B^{-+}(x)$ :

$$C_2 : f_{A1}^+(x)/f_{A2}^-(x) = f_{B1}^-(x)/f_{B2}^+(x).$$

Since the inequalities [2.1] and [2.2] hold for  $A$  and  $B$ , and because the marginal rates of substitution are decreasing,  $C_2$  is located, from the point of view of  $A$ , above  $C_1$ . The two contract curves then generate three regions:  $\Omega_1$  under  $C_1$ ;  $\Omega_2$  between  $C_1$  and  $C_2$  (both included);  $\Omega_3$  above  $C_2$ .

- $x^0 \in \Omega_1$ :

Since in  $\Omega_1$ ,  $A$ 's left-hand and  $B$ 's right-hand marginal rates of substitution are such that  $f_{A1}^-(x^0)/f_{A2}^+(x^0) < f_{B1}^+(x^0)/f_{B2}^-(x^0)$ ,  $A$  (resp.  $B$ ) can improve his anticipated utility by



performing exchanges where he supplies (resp. demands) good 1 and demands (resp. supplies) good 2. On the contrary, since  $f_{A1}^+(x^0)/f_{A2}^-(x^0) < f_{B1}^-(x^0)/f_{B2}^+(x^0)$ ,  $A$  (resp.  $B$ ) cannot improve his anticipated utility by giving up good 1 (resp. good 2) in exchange of good 2 (resp. good 1). A general equilibrium is thus a price vector  $p^*$  for which the budget line intersects  $C_1$  in  $x^*$  where

$$f_{A1}^-(x^*)/f_{A2}^+(x^*) = f_{B1}^+(x^*)/f_{B2}^-(x^*) = p_1^*/p_2^*.$$

- $x^0 \in \Omega_3$ :

This is the exact symmetric of the previous case. The respective values of  $A$ 's and  $B$ 's marginal rates of substitution are now such that  $f_{A1}^+(x^0)/f_{A2}^-(x^0) > f_{B1}^-(x^0)/f_{B2}^+(x^0)$  and  $f_{A1}^-(x^0)/f_{A2}^+(x^0) > f_{B1}^+(x^0)/f_{B2}^-(x^0)$ . Only exchanges involving (from  $A$ 's point of view) a demand of good 1 and a supply of good 2 might improve both  $A$ 's and  $B$ 's anticipated utility, equilibrium lying in  $C_2$ , so that:

$$f_{A1}^+(x^*)/f_{A2}^-(x^*) = f_{B1}^-(x^*)/f_{B2}^+(x^*) = p_1^*/p_2^*.$$

- $x^0 \in \Omega_2$ :

In  $\Omega_2$ , the marginal rates of substitution are such that  $f_{A1}^+(x^0)/f_{A2}^-(x^0) < f_{B1}^-(x^0)/f_{B2}^+(x^0)$  and  $f_{A1}^-(x^0)/f_{A2}^+(x^0) > f_{B1}^+(x^0)/f_{B2}^-(x^0)$ .  $A$ 's and  $B$ 's indifference curves relative to  $x^0$  have henceforth no other common point than  $x^0$  itself, so that no anticipated exchange would be Pareto-optimal. The contract set is then the entire region  $\Omega_2$ , within which equilibrium supply and demand are zero, at equilibrium prices belonging to a close interval defined by:

$$\sup (f_{A1}^+(x)/f_{A2}^-(x), f_{B1}^+(x)/f_{B2}^-(x)) \leq p_1^*/p_2^* \leq \inf (f_{A1}^-(x)/f_{A2}^+(x), f_{B1}^-(x)/f_{B2}^+(x)).$$

The particularity of exchange between two agents who share asymmetric sensitivity to gains and losses hence comes from the existence of a contract set – and not only of a contract curve. Equilibrium final situations of optimal trajectories are then distinct from initial situations only when the latter are located outside the contract set. And in this case, they always lie on the borders of the contract set – that is, either on  $C_1$ , or on  $C_2$ <sup>18</sup>.

---

<sup>18</sup> For instance, on figure A2,  $(x^{*E}, p^{*E})$  is an equilibrium for  $x^{0E}$ ,  $(x^{*G}, p^{*G})$  for  $x^{0G}$ , and  $(x^{*F}, p^{*F})$  for  $x^{0F}$ .

## References

- Becker, Gary S. and Stigler, G.J. (1977). De Gustibus non est Disputandum. *American Economic Review*, 67(2), march.
- Bentham, Jeremy (1828-43). *The Works of Jeremy Bentham*, Published under the Superintendance of his Executer, John Bowring, Edimburgh: Tait.
- Bentham Jeremy (1828-43) [1782-7]. *The Rationale of Reward*. In John Bowring (ed.), *The Works of Jeremy Bentham*, part. VII, 189-266.
- Bentham Jeremy (1828-43) [1785-6]. *Principles of the Civil Code*. In John Bowring (ed.), *The Works of Jeremy Bentham*, part. II, 297-364.
- Bentham, Jeremy (1948) [1789]. *An Introduction to the Principles of Morals and Legislation*, edition of 1823, New York, Hafner Press.
- Bentham, Jeremy (1828-43) [1797-8]. *Outline of a Work Entitled Pauper Management Improved*. In John Bowring (ed.), *The Works of Jeremy Bentham*, part. II, 369-439.
- Bentham, Jeremy [1801]. *Sur les Prix*, mss. Dumont [L], Bibliothèque Publique et Universitaire, Genova [manuscript].
- Bentham, Jeremy (1828-43) [1831]. *Pannomial Fragments*. In John Bowring (ed.), *The Works of Jeremy Bentham*, part. IX, 211-230.
- Black, Collison R.D. (1988). Bentham and the Political Economists of the Nineteenth Century. *Bentham Newsletter*, 12.
- Duesenberry, James S. (1949). *Income, Saving and the Theory of Consumer Behavior*, Cambridge, Harvard University Press.
- Grether, D. And Plott, C. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review*, 69(4), 623-38.
- Guidi, Marco E.L. (1993). L'utilitarisme et les origines du savoir économique moderne: la thèse de la douleur de Locke à Bentham. *Oeconomia*, 18, December.
- Friedman, Milton and Savage, Leonard J. (1948). The Utility Analysis of Choices Involving Risk. *Journal of Political Economy*, 56, August.
- Halévy, Elie (1995) [1901-4]. *La formation du radicalisme philosophique*, Paris: Presses Universitaires de France.
- Holt, Charles A. (1986): Preference Reversals and the Independence Axiom. *American Economic Review*, 76, June.
- Jevons, W. Stanley (1866) [1862]. A Brief Account of a General Mathematical Theory of Political Economy. *Journal of the Royal Statistical Society*, XXIX, June.
- Jevons W. Stanley (1878) [1871]. *The Theory of Political Economy*, 2<sup>d</sup> ed., London: Macmillan.
- Lichtenstein, Sarah and Slovic, Paul (1971). Reversal of Preferences Between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology*, 89, July.
- Loomes, G. And Sudgen, R. (1982). Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *The Economic Journal*, 92(6), 805-24.

- Lundin, Hilda G. (1920). The Influence of Jeremy Bentham on English Democratic Development. *University of Iowa Studies in the Social Sciences*, 37(7), November.
- Machina, Mark M. (1987). Choice under Uncertainty: Problems Solved and Unsolved. *Journal of Economic Perspectives*, 1(1), Summer.
- Mill, James (1992) [1818-9]. Education, *Supplement of the Encyclopaedia Britannica* [1824] in *James Mill Political Writings*, Cambridge: Cambridge University Press.
- Mill, James (1967) [1829]. *Analysis of the Phenomena of the Human Mind*, 2<sup>d</sup> ed., John Stuart Mill ed. (1869), New-York: A.M. Kelley.
- Pradel, Jean (1989). *Histoire des doctrines pénales*, Paris: Presses Universitaires de France.
- Quiggin, John (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization*, 3, 323-43.
- Requier-Desjardins, D. (1996). Temps et préférences dans la nouvelle théorie du consommateur, *Conference Time and Economics*, Toronto.
- Schumpeter, Joseph A. (1954). *History of Economic Analysis*, 2<sup>d</sup> ed., London: G. Allen & Unwin.
- Semple, Janet (1993). *Bentham's Prison : a Study of the Panopticon Penitentiary*, Oxford: Clarendon Press.
- Sigot, Nathalie (1993). "Be quiet", mais modérément: le rôle de l'Etat dans la pensée économique de Jeremy Bentham. *Revue Economique*, 44(1), January.
- Sigot, Nathalie (1995). *L'Utilitarisme Benthamien à la Rencontre de l'Economie Classique*, PhD Dissertation, Centre d'Histoire de la Pensée Economique, University of Paris I Panthéon-Sorbonne.
- Stark, William (1946). Jeremy Bentham as an Economist. II. Bentham's Influence, *Economic Journal*, 56, 583-608, December.
- Thaler, Amos and Tversky, Richard H. (1990). Anomalies: Preference Reversals, *Journal of Economic Perspectives*, 4(2), Spring.
- Weizsäcker, C.C. Von (1971). Notes on Endogeneous changes of tastes. *Journal of Economic Theory*, 3, 345-372.