



**HAL**  
open science

## Collective interest vs. individual interest in Bentham's Felicific Calculus. Questioning welfarism and fairness

Antoinette Baujard

► **To cite this version:**

Antoinette Baujard. Collective interest vs. individual interest in Bentham's Felicific Calculus. Questioning welfarism and fairness. *European Journal of the History of Economic Thought*, 2010, 17 (4), pp.607-134. 10.1080/09672567.2010.483067. halshs-00528587

**HAL Id: halshs-00528587**

**<https://shs.hal.science/halshs-00528587>**

Submitted on 18 Nov 2016

**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.

COLLECTIVE INTEREST VS. INDIVIDUAL INTEREST IN  
BENTHAM'S *FELICIFIC CALCULUS*. QUESTIONING  
WELFARISM AND FAIRNESS

Antoinette Baujard <sup>12</sup>

*Abstract.* The core idea of utilitarianism for Bentham is to establish that only individual utilities count in social welfare. There can be two distinct interpretations of this apparently simple principle. According to one view, individual utilities represent the basic information for the calculation of social welfare: this is how utilitarianism works. According to a second view, social welfare is maximized if and only if individual utilities are maximized: this is what justifies utilitarianism. This aim of this paper is to show: that these two interpretations should not be confused; that they correspond to distinct definitions of welfarism; that they are likely to conflict; and that as a consequence we can draw important and surprising conclusions for political philosophy and economic science. One such conclusion is that fairness should be prior to goodness in a consistent Benthamian doctrine.

*Short abstract.* The core idea of utilitarianism for Bentham is to establish that only individual utilities count in social welfare. There can be two distinct interpretations of this apparently simple principle. Individual utilities represent the basic information for the calculation of social welfare (how utilitarianism works). Or: social welfare is maximized if and only if individual utilities are maximized (what justifies utilitarianism). This aim of this paper is to show that these two interpretations correspond to two distinct definitions of welfarism, which are likely to conflict; and that, as a consequence, fairness should be prior to goodness in a consistent Benthamian doctrine.

*JEL classification code:* B12, B31, D63

*Key-words:* interest, utility, is and ought, external effects, goodness and fairness, welfarism, formal welfarism, ethical welfarism

---

<sup>1</sup>Antoinette Baujard, CREM, University of Caen, F-14 000, [Antoinette.Baujard@unicaen.fr](mailto:Antoinette.Baujard@unicaen.fr)

<sup>2</sup> Previous versions of this paper have benefited from the reactions of participants at the 2006 ESHET meeting in Porto (April 2006), the 3LB workshop (Strasbourg, November 2006), the Charles Gide Days (Aix-en-Provence, December 2006), the PGPPE winter '08 meeting on the History of Public Economics (Paris, December 2008) and the third JFFoS symposium (Shonan Village, Japan, January 2009). The research on which the paper is based has been supported by funds from ANR and the 3LB programme. I here thank André Lapidus, Herrade Igersheim, Keith Tribe, and two anonymous referees for their valuable comments. Needless to say, any remaining deficiencies are my own, as are the views expressed in this paper.

## INTRODUCTION

“It is vain to talk of the interest of the community, without understanding what is the interest of the individual.” So Bentham writes in the first chapter of *The introduction to the principles of morals and legislation*. It is now standard to call “welfarist” the view according to which the collective interest derives only from individual utilities.<sup>3</sup> The fact that Bentham considers the sum of individual utilities to be the criterion of collective welfare may be controversial, but it cannot be doubted that he is himself welfarist.<sup>4</sup> Having established that, new issues arise: What is this individual interest which is linked to community interest? Is it that which guides individual behaviour? Is the link between the positive and the normative aspect of interest straightforward? What would be the exact definition of Bentham’s welfarism?

The crux of all these issues lies in the opposition between particular interests and collective interests, as well as in the distinction between ‘is’ and ‘ought’. I suggest that consideration of these two oppositions will enrich our appreciation of Bentham’s writings. Focusing on his *felicific calculus*, their reformulation permits us to understand the ethical and analytical link between private and collective interests.

This article is organized as follows. In the first section I introduce some standard readings of Bentham’s view of ‘is’ and ‘ought’ and discuss relevant ways in which we can move beyond mere registration of their distinction. I propose to reformulate the problem by specifying the normative or the positive formulation of individual or collective interests. In the second section, I show that a duality of the individual ‘ought’ is derived from this analysis: among the two notions of the individual ‘ought’, one relates to the individual level, the other to the collective level. In the third section, this result is shown to be relevant both for Bentham studies and for contemporary debates in welfare economics. I will in particular claim that this view entails, in most contexts, a necessary distinction between a formal definition of welfarism (at the theoretical level), and an ethical definition of welfarism (at the philosophical level).<sup>5</sup> From this will be derived a disturbing conclusion: it will be

---

<sup>3</sup>The term was first popularized in two articles written by Amartya Sen, one published in the *Economic Journal* (Sen, 1979a), and the other one in the *Journal of Philosophy* (1979b).

<sup>4</sup>The following statement is just one of several similar passages: “De quoi se compose le bonheur total si ce n’est des utilités individuelles?” (in Bentham 1831, Vol.1: 26) [All quotations in French are written in French in the original version.]

<sup>5</sup>This paper follows Baujard (2009), entitled “A Return to Bentham’s felicific calculus I. From Moral Welfarism to Technical Non-welfarism”, in which the presentation of the calculus leads to consideration of the distinction between technical welfarism—at the applied level—and ethical welfarism—at the philosophical level.

shown that the link made between private and collective interests in the standard reading of Bentham is mistaken, a finding that also applies to utilitarianism in general. There is a consistent and inescapable consequence: that in Bentham's utilitarianism fairness should be given priority over goodness.

## 1 THE NORMATIVE AND THE POSITIVE ANALYSIS OF UTILITY

### Standard views of the 'is' and the 'ought' in Bentham's works

"I. Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. [...]

II. [...] By the principle of utility is meant that principle which approves or disapproves of every action whatsoever (...), and therefore not only of every action of a private individual, but of every measure of government." (Bentham, 1789: 1-2)

In Bentham's works, utility is the axis both of a positive principle and a normative principle.<sup>6</sup> The former is relative to 'is': seeking pleasure and avoiding pain guides individual action; psychological law or psychological hedonism is at stake. One can speak of the "measure of the pleasures of the senses".<sup>7</sup> The latter is relative to 'ought': the society's objective is to obtain the greatest happiness for the greatest number;<sup>8</sup> universal hedonism is at stake. One can speak of the "measure of a good government."<sup>9</sup> These two elements define a principle of utility, valid at the individual *and* at the collective level.<sup>10</sup>

According to some authors both classical and contemporary utilitarianism are characterized by a controversial tension between the positive and the normative aspects of utility (Halévy, 1901; 1905; Harrison, 1983; Audard, 1999). Ph. Mongin

---

<sup>6</sup>See Mongin (1995), Mongin and Sigot (1999).

<sup>7</sup>Cot (2000: 290)

<sup>8</sup>Or, rather, the principle of the greatest happiness; see Burne (1949).

<sup>9</sup>Cot (2000: 290)

<sup>10</sup>"By the principle of utility is meant that principle which approves or disapproves of every action whatsoever. According to the tendency it appears to have to augment or diminish the happiness of the party whose interest is in question: or, what is the same thing in other words to promote or to oppose that happiness. I say of every action whatsoever, and therefore not only of every action of a private individual, but of every measure of government." (Bentham 1789: 5)

(1995: 385) in particular regrets that “the two views remain alien from one another.”<sup>11</sup> Applying this same principle to two different levels, normative and positive on the one hand, individual and collective on the other hand, is thought to be problematic, and requires some clarification. While utility is the basis of what *is* and what *ought to be*, the ‘is’ and the ‘ought’ do not always coincide. The distinction might even be considered as evidence of a contradiction.<sup>12</sup> Without committing ourselves to such extreme positions, the links between the normative and the positive aspects of utilitarianism have been questioned by utilitarians and their commentators from different perspectives: the difference between psychological hedonism, and universal hedonism (Sidgwick, 1874; Guidi, 2007); the role of egoism or self-interest (Vergara, 1998; Mongin and Sigot, 1999); individual or extra-individual prudence (Halévy, 1905); the distinction between the economic and the politic domain (Halévy, 1901; Sigot, 2000); the role of State (Sigot, 1993); the individual and the collective aspect of the utility principle (Sigot, 1993, 2001); the distinction between a moral utility function and a personal utility function (Harsanyi, 1955; 1992); the issue of laundering preferences (Goodin, 1986), or, more generally, the choice of a model of utility suited to the ethical project (Broome, 1991; Haslett, 1990). Nevertheless, if all of these positions establish—or negate—the existence of a contradiction, none of them confronts the persistence of such an opposition. Beyond the mere acknowledgement of a contradiction or of a distinction, a formal differentiation between the ‘is’ and the ‘ought’ is thought to be necessary, raising the question of their controversial link. Here we intend to discuss the development of this debate concerning tensions and analytical links between the ‘is’ and the ‘ought’ by reference to utilitarian theories of justice, taking Bentham’s utilitarianism as a fundamental reference. We shall now introduce our approach to the problem, and demonstrate that it will also require consideration of the link between private and collective levels.

The distinction between the ‘is’ and the ‘ought’ acknowledges a divide that *does* exist, but which *should* not be allowed to persist. It exists because individuals, acting in accordance with their own happiness, do not always act in such a way that the

---

<sup>11</sup>“We know of no contemporary author who considers the relationship between the is and the ought within utilitarianism.” Our translation of [henceforth, *Trans.*]: “Nous ne connaissons pas d’auteur contemporain qui thématise la relation de l’être et du devoir-être à l’intérieur de l’utilitarisme.(...)”

<sup>12</sup>See for instance Harrison (1977: 654): “Mr. Parekh also claims, rather darkly (p.116), that Bentham gave no reason for holding that ‘personal and political life should both be regulated by the same principle’.” See also the position of Friedman (1948) who points out Bentham’s “refusal to see the reality of conflict between the individual interest and that of the community.”

collective interest<sup>13</sup> is maximised. The ‘is’ and the ‘ought’ may therefore be in conflict,<sup>14</sup> yet imply no inconsistency in Bentham’s utilitarianism (see Guidi, 2004). As a consequence, the community’s interest would be at a higher level than that of the status quo if a different set of incentives could be designed such that each member would, eventually, act in such a way that the higher interest is realized.<sup>15</sup> Bentham seeks to generate a transparent and rational society, organized through both the principle of utility defined at the collective level, and the principle of individual interest (See Cot (2000)).

### **Individual interests and collective interest in Bentham’s works**

The problem of the distinction between the is and the ought is therefore shifting towards the reconciliation of individual interest with the community interest. Whenever they coincide this gap is bridged. According to Halévy’s reading of utilitarianism, there are three ways of identifying individual and collective interests. The first is the sympathetic fusion of interests; it supposes that individual interest is not selfish interest. The second concerns the natural identification of interests: if every one were to spontaneously act in an ethic manner—not necessarily consciously—there would in fact be no conflict between individual and collective interests. This apparently trivial statement comes from Bentham’s economic philosophy. The resulting spontaneous order finds echoes in libertarian theory, where government intervention is neither necessary, nor expected. These two approaches are however insufficient. Bentham devotes the principal part of his work to developing the third approach embedded in his philosophy of law: the artificial identification of interests. Outside intervention is necessary if collective and

---

<sup>13</sup>I will not here discuss the difficulties of determining what “interest” is in actual situations. See Laval (1993). I will employ a minimalist definition of “interest”, as what is good for the individual or the community.

<sup>14</sup> “But does not the moralist consider himself necessarily condemned to resolve this absurd problem: in a society in which each individual is necessarily selfish, how can one nonetheless render all disinterested?” *Trans.* “Mais alors le moraliste ne se voit-il pas condamné, par profession, à résoudre ce problème absurde : dans une société où chaque individu est égoïste, et se doit à lui-même de l’être, obtenir cependant que tous les individus soient désintéressés ?” Halévy (1905, III: 197)

<sup>15</sup>“Homo oeconomicus is not generally motivated to make the right choices from an aggregate perspective. Because most choices have externalities, the optimal decision for an individual is often sub-optimal for society as a whole. Much of Bentham’s work was devoted to the design of ‘political sanctions’ meant to change the balance of an agent’s pleasure and pain via the police power of the State, but he also acknowledged the complementary role of ‘moral sanctions’ in internalizing externalities, and thus in inducing the right choices for promoting aggregate happiness.” Warke (2000a: 374)

individual interests are to achieve compatibility.<sup>16</sup> Such intervention takes the form either of external intervention—the Civil or Penal code, the legal-administrative supervision of the town or city, or suitable legislation (Bentham, 1830; 1827; 1872); or internal intervention, resorting to a deontologist (Bentham, 1831).

There are two ways of tackling the problem of identification of interests in Bentham's work. The first consists in considering the diversity of the modalities of identification. Some authors have noted an opposition between his economic and political stances, defending a natural identification in the former case and an artificial identification of interest in the latter.<sup>17</sup> But it is difficult to uphold this schizophrenic relationship between the economic and the legislative domain. Other writers have recently shown

---

<sup>16</sup> "If all men wish to be happy, but cannot be so, is it not because the means which individuals use to this end are mutually contradictory? [...] Since this reduces the attraction of pleasure, it is necessary to threaten the infliction of pain equal at least in intensity to the pleasure to which the individual aspires. Such threats convert these actions into offences. The science of intimidation, that is, legislation, general utility, these are the motivation, and pain is the sanction of the obligations which it imposes. In society, the Legislator is the great dispenser of pleasure and pain. It is he who creates moral order, the equilibrium of interests. Society is the construct of his artifice. There we find his application of what we have called the principle of the artificial identification of interests." *Trans.* "Tous les hommes veulent être heureux ; mais ne peut-il pas, ne doit-il pas arriver que les moyens employés par les divers individus pour être heureux soient contradictoires entre eux ? [...] Puisque celui-ci cède à l'attrait du plaisir, il faut le menacer de l'infliction d'une douleur au moins égale en intensité au plaisir auquel il aspire. Par ces menaces, des actes sont érigés en délits. La science de l'intimidation, voilà la législation ; l'utilité générale est la raison d'être, et la peine est la sanction des obligations qu'elle impose. [...] Le législateur est, dans la société, le grand dispensateur des plaisirs et des peines. C'est lui qui crée l'ordre moral, l'équilibre des intérêts. La société est l'œuvre de ses artifices. – Ainsi trouve son application ce que nous avons appelé le principe de l'identification artificielle des intérêts." Halévy (1905, III: 216-217)

<sup>17</sup> "However, the two principles upon which rest both the legal and the economic philosophy of the Benthamites are contradictory. This contradiction entirely shatters the familiar precepts of Benthamism. Do we then have to resolve this contradiction by saying that each of the two principles has its application in distinct domains that the principle of the artificial identification of interests is the true principle of the science of law, and the principle of natural identity of interests the true principle of economic science? It is obvious that Bentham has borrowed from two contradictory sources in creating his system." *Trans.* "Or, les deux principes sur lesquels reposent respectivement la philosophie juridique et la philosophie économique des Benthamites sont deux principes contradictoires : la contradiction éclate à chaque instant dans les formules courantes du Benthamisme. [...] Faudra-t-il donc se borner à résoudre la contradiction en disant que les deux principes trouvent chacun son application dans un domaine distinct, que le principe de l'identification artificielle des intérêts est le véritable principe de la science du droit et le principe de l'identité naturelle des intérêts le véritable principe de la science économique ? Il est évident que Bentham a emprunté à deux sources contradictoires d'un même système." Halévy (1905, III: 219)

that Bentham also supports state intervention in the economic domain.<sup>18</sup> An exception to this demarcation moderates this view.

The second approach to consideration of the gap between individual and collective interests is to focus on the case of artificial identification of interests. I therefore focus upon those states in which artificial identification is the only reliable modality of a junction of interests. The same conclusion can be obtained: according to the principle of utility, any domain, political or economic, is likely to need—or is not likely to need—a degree of state intervention<sup>19</sup>. As a matter of fact, Bentham introduces an economic means for managing judicial, political and non-market issues.<sup>20</sup> The tension between the ‘is’ and the ‘ought’ is not just a problem of adjudication; Bentham proposes an economic means of dealing with it. Deeper understanding of the character of the artificial identification of interests is therefore the most fitting way to analyze the issue of the linkage between the is and the ought.

We direct our attention to those specific cases where an artificial identification of interests is required: how to ensure that individual actions, arising from decisions made at the individual level, might respect the principle of utility defined at the collective level. It should be recalled that the positive principle applies at the individual level, while the normative principle applies at the collective level. We hence have to deal with two issues: the positive vs. normative problem on the one hand, and individual interests vs. collective interest on the other. We contend that simply stating this distinction is likely to be merely confusing, and that a clear statement of the link between each of the levels is necessary.

Even though, to be meaningful, the normative principle requires an aggregation stage, both the positive and the normative principles rely on individual interests and hence on the utility of the things or actions leading to this interest. If there exists a tension between the hedonic and the ethical principles, between individual interests and the collective interest, it should be clearly enough revealed *via* the process of defining individual utility.

### **A reformulation of the ‘is’ and ‘ought’ issue**

---

<sup>18</sup>See Sigot (1993) among others.

<sup>19</sup>On this see Sigot (1993; 2001).

<sup>20</sup>“It should not be forgotten that political economy was for Bentham a branch of the science of legislation” Guidi (2002: 176). In this perspective, Bentham created the “pre-history” of welfare economics and of the economic analysis of law.

This movement from 'is' to 'ought' and back, between the individual and the collective level, is constantly present in the texts devoted to Bentham's "felicific calculus".<sup>21</sup> To clarify my argument I will first introduce some notation, as summarized in table 1, to be used in the following discussion.

Table 1: Behaviours and interests guiding behaviours in Bentham's utility calculus

	Positive principle Description of behaviours	Normative principle Evaluation of the situations
Individual level	$\alpha_i$	$U_i / V_i$
Collective level	$\alpha = (\alpha_1, \dots, \alpha_i, \dots, \alpha_n)$	W

Let us first present the positive part. The individual level is described by identifying actions  $\alpha_i$  for individual  $i$ . At the collective level, the positive analysis hence consists in the combination of the list of actions of all of the  $n$  individuals  $(\alpha_1, \dots, \alpha_i, \dots, \alpha_n)$ . This could be considered to be a mere description of social states. Actions or social states are, in any case, justified or prompted by some normative principles. At the individual level, each individual seeks her own interest<sup>22</sup> or happiness, represented by the functions  $U_i$  or  $V_i$ , as we shall see below. Her motive to act is the expectation of more pleasure and less pain. Utility is the property of things or actions to provide more pleasure and to limit pain; it is therefore the guiding principle for individual behaviours. This relation between motives and interest for one person  $i$  from the community generates the choice of action  $\alpha_i$ , (as we described it) at the positive

---

<sup>21</sup>See the references to different treatments of the utility calculus in Mitchell (1918: 164). See also Guidi (2007).

<sup>22</sup>See Mongin and Sigot (1999) and references quoted in their article.

level.<sup>23</sup> Hence, as  $\alpha_i$  is maximizing  $i$ 's interest, the list of individual choices  $(\alpha_1, \dots, \alpha_i, \dots, \alpha_n)$  is supposed to maximize each individual's interest. At the collective level, what ought to be is the greatest happiness (of the greatest number). The principle of utility is defined in the following way: each action ought to be assessed by this principle, i.e. an action should be completed if it does lead to more pleasure and less pain for the community; it should not be done in the converse case. Let us call this objective  $\mathcal{W}$  for Social Welfare.

Let us now deepen the relationship between individual utilities and Social Welfare. Consistently in table 1,  $(\alpha_1, \dots, \alpha_i, \dots, \alpha_n)$ , and  $\mathcal{W}$  may be considered as the resulting level of individual utility or collective welfare. The following discussion should aid thorough description of the underlying functions, in particular showing the stakes involved by the domain of individual utility functions. The principle of utility defines Social Welfare as a function of individual interests or happiness. There are some fragments of textual evidence, though very few, in which Bentham defends the idea of a *sum* of individual interests, attributing the same weight to each. In his own words, “the interest of the community is one of the most general expressions that can occur in the phraseology of morals: no wonder that the meaning of it is often lost. When it has a meaning, it is this. The community is a fictitious *body*, composed of the individual persons who are considered as constituting as it were its *members*. The interest of the community then is, what? —the sum of the interests of the several members who compose it.” (Bentham, 1789: 5). More generally, Bentham's utilitarianism requires that this social welfare should be based exclusively on individual utilities: note that this is specifically the definition of welfarism. It is not however obvious in Bentham's texts whether individual actions should eventually be justified at the individual level, or by the utility principle at the collective level. On the first interpretation, person  $i$  would justify her action  $\alpha_i$  because it maximizes her utility  $U_i$ , without further consideration. On the second interpretation, the same action  $\alpha_i$  would be justified because it contributes to maximize social welfare  $\mathcal{W}$ . Some could claim that this probably amounts to the same principle: as  $\mathcal{W}$  is based on each individual's utilities among which  $U_i$  and  $\alpha_i$  maximizes  $U_i$ , each action eventually contributes to the increase of social welfare  $\mathcal{W}$ . This is certainly so in some cases. This would notably be the case if we supposed complete independence between individuals.<sup>24</sup> But this miraculous conjunction obviously fails to be true in the specific

---

<sup>23</sup>Note that there can be different motives as explanation of the same behaviour; each  $\alpha_i$  is then the result of a function of some complex—and, probably, multi-dimensional (See Warke (2000b))—combination of feeling or anticipation of pain or pleasure.

<sup>24</sup>The action of Miss  $i$  would be independent of the action of Mr.  $j$  if she never once changed her mind once she knew what he had done. If she is gregarious, she might want to change the colour of

cases where the artificial junction of interests is needed, i.e. as soon as we account for interactions and we consider the effect of the passage of time. Then the principles defined at the individual or at the collective level do not merge into one.

As a consequence, the normative principle for an individual might be different, depending on whether we focus upon individual or collective interests. In the former case, what is at stake is really the individual interest ( $U_i$ ) in itself<sup>25</sup>. In the latter case, what is important is eventually how mister  $i$ 's individual interest ( $V_i$ ) contributes to the greatest happiness of the greatest number. The rest of this paper aims at determining whether or not there is any difference between these two versions of the individual objective, called ( $U_i$ ) and ( $V_i$ ) within Bentham's doctrine, and, if so, the meaning and the consequences of such a difference. So that we might establish what analytical links there are between the positive and the normative side it is necessary to demonstrate the existing or desired links between  $\alpha_i$ ,  $(\alpha_1, \dots, \alpha_i, \dots, \alpha_n)$ ,  $W$ ,  $U_i$  and  $V_i$ .

## 2 TWO NOTIONS OF INDIVIDUAL INTERESTS

### The design of distinct incentives for different individual interests

A behaviour  $\alpha_i$  can be judged according to normative criteria, but not the motives explaining this behaviour.<sup>26</sup> The principle underlying behaviour is self-interest, which is justified by the definition of the welfarist principle. An appropriate choice of each individual behaviour  $\alpha_i$  is hence supposed to maximise each personal  $U_i$ —at least in the event of a correct calculation. At the collective level,  $(\alpha_1, \dots, \alpha_i, \dots, \alpha_n)$  is the description of all individual behaviours.  $W$  should evaluate the consequence of this list, hence of the complex interactions between individual actions. Bentham recognizes there are cases for which the combination of these decentralized primary individual behaviours  $(\alpha_1, \dots, \alpha_i, \dots, \alpha_n)$  do not induce spontaneously the best social

---

her blouse, for example, so that she is wearing the same colour as he is. These kinds of preferences should not be excluded from the present study.

<sup>25</sup>“3. It is said to be a man's interest that the act, the event, or the state of things in questions should have place, in so far as it is supposed that—upon and in consequence of its having place—good, to a greater value, will be possessed by him than in the contrary case. In the former case, interest corresponds to a single item in the account of good and evil; in the latter case, it corresponds to a balance on the side of good.” Bentham (1817: 208)

<sup>26</sup>“There is not any such thing as a bad motive. [nor good ones].” Bentham (1817: 215)

optimum  $\mathcal{W}$ .<sup>27</sup> If the social optimum—the greatest  $\mathcal{W}$ —cannot be reached spontaneously, one should design different incentives, so that seeking her individual interest now induce individuals to act ( $\alpha'_i$ ) in the right direction, i.e. so that ( $\alpha'_1, \dots, \alpha'_i, \dots, \alpha'_n$ ) maximizes  $\mathcal{W}$ .

The aim is to obtain a specific list of behaviours ( $\alpha'_1, \dots, \alpha'_i, \dots, \alpha'_n$ ), different from the previous list ( $\alpha_1, \dots, \alpha_i, \dots, \alpha_n$ ), and which is now likely to maximise  $\mathcal{W}$ . There is no inner contradiction in this description. Nevertheless, this reasoning is specific to the collective level of analysis: we talk of collective interest ( $\mathcal{W}$ ), assessing the result of all individual behaviours together. But the actual policies are to be enforced at the individual level: the new system of individual incentives is designed so that the new individual action  $\alpha'_i$  is distinct from  $\alpha_i$ , given that individuals keep on seeking their individual interest. This simply means that this design has recomputed the actual individual interest in some utility  $V_i$ , indeed distinct from  $U_i$ .

Let us summarize. As long as all individuals seek their own interest ( $U_i$ ) <sub>$i \in N$</sub> , they do not necessarily seek the greatest collective interest  $\mathcal{W}$ . If a new system of incentives manages to reorientate individual interests into ( $V_i$ ) <sub>$i \in N$</sub> , which is likely to be different from the previous list ( $U_i$ ) <sub>$i \in N$</sub> , then individuals seeking their new interests ( $V_i$ ) <sub>$i \in N$</sub>  will eventually maximise  $\mathcal{W}$ . In other words, there exist two kinds of individual interest for Bentham. We will now call  $U_i$  the self-defined interest for individual  $i$ ; and  $V_i$  the collectivity-defined interest for individual  $i$ . Self-defined interest does not strictly imply egoism, but a restriction of relevant information for defining one's individual interest. Collectivity-defined interest does not imply altruism, but an extension of admissible information to collective states for defining one's individual interest.

As a preliminary conclusion, we claim that, if no contradiction lies between the 'is' and the 'ought' in general, the key issue lies rather in the ambiguity over the individual 'ought'. When they merge, an analytical continuity between the individual and the collective level should hold. When they do not do so spontaneously, continuity is only restored by the collectively-defined interest  $V_i$ . The definition and the status of the latter hence ought to be justified since it eventually derives from a technological relation between individual behaviours. If the ethical link between individual and collective interests appears clearly in Bentham's doctrine, the whole theory can be consistent only if the positive link between them is also made explicit.

### **Ethical link between individual and collective interests**

---

<sup>27</sup>Individual behaviours can be evaluated as generating good or bad social results and we may wish to change them, but according to Bentham's doctrine we should not evaluate the motives of all individuals leading to this result as long as they derive from a desire to raise individual utility.

The ethical link between the individual interest and the collective interest established by Bentham respects the two following criteria: the principle of valuing individual interests for themselves and the rejection of natural law and fundamental rights (See Cot 1993). It is, firstly, based on the following condition: individual interests are the only elements from which collective welfare can be derived—this defines Bentham’s ‘welfarism’.<sup>28</sup> Bentham establishes that what is of interest for an individual should necessarily be of interest for the community. If an action is likely to increase somebody’s interest, it should increase the interest of the entire community. This basically corresponds to the weak Pareto criterion founding the social welfare function. Trouble arises when (among other things) an increase in somebody’s interest is associated with a decrease in the interest of some other member of the community. If one is to avoid impossibilities or incompleteness, the fairness basis of the trade-off has to be clarified.

Secondly, the collective interest is indeed built by some specific combination of individual interests. The aggregation could be a simple summation of individual interests,<sup>29</sup> but this option is sometimes controverted (See Shackleton 1972; Goldworth 1979). It could also be any version of possible aggregation of individual interests.<sup>30</sup> We can note that relying on a sum implies acceptance—or even justification—of a reduction in one person’s level of private interest insofar as it permits the overall improvement of the collective interest. At first sight, this can be considered to be a contradiction with the primary principle of welfarism, since it implies the sacrifice of certain interests in favour of the community, rather than defining the community interest in respect to and according to the evaluations of individual interests.

This doubt gives rise to a third remark. The question of compatibility of individual interests is translated into a problem of weighting the pains of some with the pleasure of others, and the gap between the different nature of individual and collective interest is not easily bridged by the simple assumption of aggregation. Welfarism requires that individual interests be at stake at the collective level, even though some may be sacrificed to reach a higher level of collective welfare. The identification of

---

<sup>28</sup>“It is vain to talk of the interest of the community without understanding what is the interest of the individuals.” (Bentham, 1789: 3); *Trans.* “De quoi se compose le bonheur total si ce n’est des unités individuelles ?” (Bentham (1831a: 26); on utilitarianism and welfarism, see Sen (1979b).

<sup>29</sup>For instance: “By *the universal interest*, understand the aggregate of those same interests” (Bentham 1823: 269); “The interest of the community then is, what? —the sum of the interests of the several members who compose it.”(Bentham 1789: 4)

<sup>30</sup>For a simple consideration of the majority of persons, see Ayer’s view in Keeton and Schwartzberger (1948), discussed by Mundle (1949: 120).

the relevant individual interests is therefore questionable, and we will see how this leads into a discussion over whether these are better defined by  $V_i$  or  $(U_i)$ .

### **Positive link**

Myrdal (1932: 51) claims that belief in a link between the positive and the normative level requires an assumption of natural harmony of interests: “the whole doctrine of harmony is implied in [the] logical transition from the psychological to the ethical theory”, from individual to collective ends (See Hume (1969: 299)). For Myrdal, this thesis of natural harmony is highly controversial—and, further, inconsistent with Bentham’s ambition to do away with natural law (See Cot (1993)).

Why does behaviour  $\alpha_i$ , which has been chosen to obtain the greatest  $U_i$ , not spontaneously contribute to a greater  $W$ —meaning this behaviour does not maximise the corresponding  $V_i$ , which does maximise  $W$ ? Three types of reason may explain why individual behaviour does not necessarily serve one’s own interest. First, individuals may be mistaken about their own interests because they are not well informed. Second, the ability of individuals to properly calculate are limited: they may make mistakes about the best possible choices even though they wish to seek their own interest, and even though they wish to seek the collective interest. As for the third type of reason, there exist ‘external effects’. The latter are taken into account in Bentham’s *felicific calculus* through the criteria of fecundity, purity and extension.<sup>31</sup> He also acknowledges the role of social interaction in people’s happiness.<sup>32</sup> In other words, interaction, whatever the context, is indeed likely to induce sub-optimality, as soon as private costs/benefits are different to public costs/benefits.

### **Mechanisms to enforce the link**

There is no reason why  $U_i$  and  $V_i$  should be the same, but their result *should* become the same. The entire task which Bentham has set himself consists in finding ways to *enforce this positive link* between these two different notions of individual interest.

---

<sup>31</sup>See Bentham (1789: 30). For a reconstruction of this calculus, see Baujard (2009).

<sup>32</sup> “A greater part of the pleasures of a man is subject to the will of others, and such pleasures can be had only through their agreement and co-operation. It is not possible to neglect the happiness of others without risking our own. We are not able to avoid the pain that it is in the power of others to inflict upon us if it does not accommodate their goodwill. Each man is united with the human race by the strongest of connections, that of personal interest.” *Trans.* “Une grande partie des plaisirs d’un homme est subordonnée à la volonté des autres, et il ne peut les posséder qu’avec leur concours et leur coopération. Il nous est impossible de négliger le bonheur des autres sans risquer le nôtre. Nous ne pouvons éviter les peines qu’il est au pouvoir des autres de nous infliger, si ce n’est en nous conciliant

In a Christian society, one could consider increasing the sense of duty in individuals. Yet a sense of duty is not useful, and not even necessarily efficient.<sup>33</sup> The intention which they imply is of little interest.<sup>34</sup> All that matters is actual effects. And the effects of all individuals actions, moved by the desire of more pleasure and less pain, should eventually lead to the optimal society, where the greatest happiness is obtained. Explaining  $\alpha_i$  or saying that individual  $i$  seeks his own interest  $U_i$  indeed amount to the same thing, which is the axiomatic basis of Bentham's works:<sup>35</sup> one cannot expect the normative objective to be enforced through simple means—such as laws or sense of duty—because the positive description of utility will always be more powerful. Therefore, the only way to re-direct the behaviours of individuals along the right lines is to modify their perception of their individual interest from  $U_i$  to  $V_i$ . This new perception can be obtained by making available more information about the situation, or by modifying the context for the actual calculation of individual interest.

In each case, individuals are not the best judges of what is good for them: they are mistaken in their calculus. The appeal to a deontologist<sup>36</sup> may help each individual to account for the three reasons given above in his calculus. But the deontologist is not just there to teach how to make good calculations, he also has a role in transferring information from one person to another and so building well-founded expectations of other people's behaviour.<sup>37</sup>

---

leur bon vouloir. Chaque homme est uni à la race humaine par le plus fort de tous les liens, celui de l'intérêt personnel." Bentham (1831: 158)

<sup>33</sup> "A man, a moralist, boastfully ensconced in his armchair, digresses in pompous phrases on duties and obligations. Why doesn't anyone listen to him? Because when he talks of duty, each thinks of self-interest. It is in man's nature to think of self-interest before anything else, and it is for this reason that every enlightened moralist will consider it in his interest to begin—it would be well said and well done – with duty always taking second place to self-interest." *Trans.* "Un homme, un moraliste, s'étale gravement dans son fauteuil, et là, vous le voyez dogmatiser en phrases pompeuses sur le devoir et les devoirs. Pourquoi personne ne l'écoute-t-il ? Parce que, tandis qu'il parle de devoirs, chacun pense aux intérêts. Il est de la nature de l'homme de penser avant tout à ses intérêts, et c'est par là que tout moraliste éclairé jugera qu'il est de son intérêt de commencer; il aura beau dire et beau faire, à l'intérêt le devoir cèdera toujours le pas." Bentham (1831a: 17)

<sup>34</sup>See Bentham (1817: 215–218).

<sup>35</sup>"Is it susceptible of any proof? It should seem not: for that which is used to prove everything else, cannot itself be proved: a chain of proofs must have their commencement somewhere. To give such proof is as impossible as it is needless." Bentham (1789: 4)

<sup>36</sup>See Bentham (1831).

<sup>37</sup>This is especially clear in the case of a coordination game.

Public opinion<sup>38</sup> and, more generally, popular sanctions,<sup>39</sup> also enforce the link between individual and collective interests, since the eventual aim is to have the impression of considering the collective interest ( $W$ ) as the objective of one's own  $V_i$ . The panopticon<sup>40</sup> is the extreme solution: using the power of popular sanction to incorporate the collective interest in each person.

Last but not least, the government modifies the system of utilities: with a system of sanctions and rewards,<sup>41</sup> individuals will now do something they ought to do, rather than what they would have done spontaneously.<sup>42</sup>

In each solution reconciling the private and the collective interest, Bentham had to exert some pressure on the definition of the individual interest for it to fit properly. The tension that exists between the spontaneous  $U_i$  and the desired  $V_i$  raises a serious issue in Bentham's doctrine.

### 3 THE CONSEQUENCES OF THE DISTINCTION

#### Two notions of welfarism in Bentham's doctrine

Bentham describes himself as what we would now call, following Sen's 1979's paper, a welfarist: "What is good for the individual is what is good for the community".<sup>43</sup>

---

<sup>38</sup>See Bentham (1831a: 30).

<sup>39</sup>See Bentham (1831a: 119).

<sup>40</sup>See Bentham (1872, Vol.4).

<sup>41</sup>See Bentham (1872), especially Vol.1 and 2.

<sup>42</sup> "The morality of utilitarians is their economic psychology as an imperative. Two centuries earlier, Hobbes had founded an entire system of social despotism upon a doctrine of utility; in fact the principle of the artificial identity of interests upon which Bentham founded his legal theory justifies the following interpretation of utilitarianism: it is the threat of chastisement inflicted by the sovereign that makes the individual associate interest with duty." *Trans.* "La morale des utilitaires, c'est leur psychologie économique mise à l'impératif. Deux siècles plus tôt, Hobbes avait fondé sur la doctrine de l'utilité tout un système de despotisme social; en fait, le principe de l'identité artificielle des intérêts, sur lequel Bentham asseyait sa théorie juridique, justifiait une telle interprétation de l'utilitarisme: c'est la menace d'un châtement infligé par le souverain qui fait pour l'individu la liaison de l'intérêt avec le devoir." Halévy (1905, III: 206)

<sup>43</sup>"[T]he greatest happiness of all the several members of the community in question, taken together, is the end to which it is my desire to see all the arrangements employed in the delineation of it directed. [...] Should it be asked, *Why* is it your desire that the greatest happiness of all the several members of the community in question should be the end to which all the several arrangements employed in the delineation of the form of government, by which that same community is governed, should be

Armed with the distinction between the two notions of individual interests,  $U_i$  and  $V_i$ , we now show that there are two distinct interpretations of this apparently simple principle of welfarism. According to one interpretation, the basic information needed for the calculation of social welfare is individual utilities; we will henceforth call this interpretation the formal definition of welfarism. According to a second interpretation, social welfare is maximized if and only if individual utilities are maximized: this is what justifies utilitarianism. We will henceforth call this interpretation the ethical definition of welfarism.

The formal definition of welfarism describes the way utilitarianism works: what is good for individual (premise) is what is good for the community (conclusion). Aggregation of individual interests, as presented above, suffices to bring this about.

Sometimes a deontologist is needed to help individuals to understand what their “true” interest is.<sup>44</sup> How should we define “true individual interest” as against mere individual interest? Consistent with the first and second reasons why collective and individual interest does not converge, the deontologist will assist individuals in recognizing their own interest whenever there is problem of information, or lack of ability to calculate. The third problem emphasised above was that of external effects: other individuals sometimes play a role in any given individual interests. Let us imagine a case in which the action of Mr. 2 has some effect upon Miss 1’s self-interest. His action was driven by his own motives: to obtain the greatest pleasure and the least pain. All the same, by taking into account the effect of Miss 2’s own action he may find that the best action in regard to his interest should have been different. This sequence calls for improved calculation, using more information and taking into account clear and reliable expectations of other people’s actions.<sup>45</sup> In this particular instance the deontologist may help each individual to become more rational, to seek their best interest, society consequently attaining a higher  $W$ . These two first types of reason why individual behaviour does not always contribute to private or collective interests do not raise any theoretical problems likely to destabilize utilitarianism.

---

directed? my answer is, – because on the occasion in question, such is the form, the establishment of which would in the highest degree be contributory to my own greatest happiness.” Bentham (1827: 7)

<sup>44</sup> “The task of the enlightened moralist is to demonstrate that an immoral act represents a false calculation of personal interest, and that a vicious man effects a faulty estimation of pleasure and pain.” *Trans.* “La tâche du moraliste éclairé est de démontrer qu’un acte immoral est un faux calcul de l’intérêt personnel, et que, l’homme vicieux fait une estimation erronée des plaisirs et des peines.” Bentham (1831a: 19-20)

<sup>45</sup> This involves strategic behaviour capable of game theoretic formalization. This specific case is covered by a coordination game, or more generally any non-zero-sum game. In each case the example needs describing more specifically, but the overall idea remains the same.

Hence the welfarist proposition is secure; it is simply a matter of information and calculation. But this might not always be true in the third type of reason.

Hence we now turn to consider external effects or, more generally, the effects of interactions upon the distinction between the two interpretations of welfarism: between formal and ethical welfarism. Imagine a case when the overly bad consequences of Miss 2's action will never have consequences for her individual interest. Bentham gives examples of this: "If such were the condition of human beings that the happiness of no one being came in competition with that of any other,—that is to say, if the happiness of each, or of any one, could receive increase to an unlimited amount, without having the effect of producing decrease in the happiness of any other, then the above expression might serve without limitation or explanation. But on every occasion, the happiness of every individual is liable to come into competition with the happiness of every other. If for example, in a house containing two individuals, for the space of a month, there be a supply of food barely sufficient to continue for that time; not merely the happiness of each, but the existence of each, stands in competition with, and is incompatible with the existence of the other."<sup>46</sup> This externality cannot be internalized by improved calculation. There is no reason why the individual should consider it as an objective in itself. Neither is there any reason why the deontologist should intervene on a strict welfarist basis if we retain the first interpretation. Yet, at the collective level, some individual interests should be traded-off against those of another. There is therefore for the society a purpose, an end, justifying the manner in which this trade-off could be made. The aggregation principle is needed to move beyond a mere clash of interests, but it is not part of the welfarist principle (See Sen (1979)). Once defined, this society's objective should become an authoritative and indirect end for each individual. Normative individual interests should therefore, for a Benthamite doctrine, be better represented after this reconstruction. In the examples analyzed above, some notion of individual interest in the integration of external effects generated should be introduced, even though the individual does not perceive them.<sup>47</sup> This eventually implies modification

---

<sup>46</sup>Bentham (1827: 6)

<sup>47</sup> "If as above, so it be, that in the situation of a ruler, whatsoever that situation be, the conduct of no man can reasonably be expected to be governed by any interest that stands, at that same moment, in opposition to that which, in his conception, is his own individual interest, it follows, that for causing it to take that direction, in which it will be subservient to the universal interest, the nature of the case affords no other method, than that which consists in the bringing of the particular interest of rulers into accordance with the universal interest.

Here, then, we have a third principle of the first rank, in addition to the two former ones. Call it, the *means-prescribing*, or *junction-of-interests-prescribing*, principle. The first declares, what *ought to be*, the next, what *is*, the last, the *means* of bringing what is into accordance with what ought to be. Meantime, this

of the definition of individual interest. At the community level, the individual ‘ought’ should be different to that at the individual level: it is imposed at the level of the community. We have therefore shown the actual distinction between  $(U_i)$  and  $(V_i)$ , and, further, the primary importance of  $(V_i)$  over  $(U_i)$ .

Henceforth, rather than asserting that ‘what is good for the individual (premise) is good for the community (conclusion)’, we claim that instead Bentham maintains the following principle: ‘What is good for the community (premise) is what is good for the individual (conclusion)’. This corresponds to the ethical definition of welfarism (rather than to the formal): if social welfare is maximized, then so are individual utilities; this inversion of causality between private and collective interests suggests that the first step is the maximisation of  $(V_i)$ , rather than of  $(U_i)$ . To be consistent with the theory of collective welfare this formulation of the principle of Bentham’s utilitarianism requires either a further criterion of aggregation and trade-off between individual utilities, or a reconstruction of individual interests.

By considering another case we will now show that the two definitions of welfarism may clash. The action of Mr. 2 has some effect upon Miss 1’s self-interest. Taking into account the effect of Mr. 2’s action on Miss 1’s action, Mr. 2 may find that the best action for his interest should have been different. In fact this can be elaborated by applying to Bentham’s framework the now familiar scenario of the prisoner’s dilemma. If one individual does pay attention to the action of another, he should be tempted to choose his dominant strategy. If both individuals carefully compare the utility of all alternatives, given the choice of the other, they will both choose their best strategy, i.e. the one leading to more pleasure and less pain, given the other individual’s behaviour. From a prisoner’s dilemma they eventually arrive at a Pareto-dominated equilibrium: a state different from the state derived from rational decentralized decisions would be better for both of them. Bentham would obviously not dispute this actual equilibrium. He proposed legislation, a constitution<sup>48</sup> and state

---

junction of interests, how can it be affected? The nature of the case admits but of one method, which is, the destroying the influence and effect of whatever sinister interest the situation of the individual may expose him in the action of; this being accomplished, he will thereby be virtually divested of all such sinister interest; remains as the only interest whereby his conduct can be determined, his right and proper interest, that interest which consists in the share he has in the universal interest, which is the same thing as to say, that interest, which is in accordance with the universal interest, taken in the aggregate.” Bentham (1827: 6–7)

<sup>48</sup>“This Constitution has for its general end view the greatest happiness of the greatest number; namely of the members of this political state: in other words, the promoting or advancement of their interests. By *the universal interest*, understand the aggregate of those same interests. This is the all-comprehensive end, to the accomplishment of which, the several arrangements contained in the ensuing code are all of them directed.” And in a note: “\* If the nature of the case admitted the possibility of any such result, the endeavour of this constitution would be—on each occasion, to maximize the felicity of every

intervention so that such a loss of collective interest might be avoided—taking into account the disutility induced by state-intervention,<sup>49</sup> or any devices leading individuals to adopt the collective interest as their own.<sup>50</sup> Intervention will become acceptable where intervention produces globally more pleasure and less pain than would result in the absence of such intervention. Bentham seeks to induce different behaviours for each individual, leading to the greatest happiness.<sup>51</sup> In the prisoner's dilemma, the mere consideration of each of the  $N$  distinct self-defined interests  $U_i$  mechanically leads to a sub-optimal social outcome; all individual utilities should clearly be transformed from  $(U_i)_{i \in N}$  to  $(V_i)_{i \in N}$  so that social welfare may be maximized. This demonstrates once again the direct connection between the second definition of welfarism (the ethical definition) and the distinction between the two notions of individual interests. Moreover, the first definition of welfarism (the formal definition) is thus not always consistent with the aim of maximizing individual utilities. In other words, formal welfarism and ethical welfarism may conflict.

Welfarism seems a nice word to describe a simple concept, upon which utilitarianism can then be founded. More thorough analysis has suddenly upset this comfortable

---

*one* of the individuals, of whose interests the universal interest is composed; on which supposition, the greatest happiness of all, not the greatest number only, would be the end aimed at.

But such universality is not possible. For neither in the augmentation given to the gross amount of *felicity*, can all the individuals in question ever be included; nor can the *infelicity*, in which the *expense* consists, be so disposed of, as to be borne in equal amount by all: in particular, such part of that same expense, as consists in the suffering produced by *punishment*.

Thus it is, that to provide for the greatest felicity of the greatest number, is the utmost that can be done towards the maximization of universal national felicity, in so far as depends on government.” Bentham (1823: 269)

<sup>49</sup> “Government cannot be exercised without coercion; nor coercion, without producing unhappiness. Of the happiness produced by government, the net amount will be—what remains of happiness, deduction made of the happiness.” Bentham (1823: 269)

<sup>50</sup> Such as the Panopticon.

<sup>51</sup> “When I say the greatest happiness of the whole community, ought to be an end of object of pursuit, in every branch of the law—of the political rule of action, and of the constitutional branch in particular, what is it that I express? —this and no more, namely that it is my wish, my desire, to see it taken for such, by those who, in the community in question, are actually in possession of the powers of government; taken for such, on the occasion of every arrangement made by them in the exercise of such their powers, so that their endeavours shall be, to render such their cause of action contributory to the obtainment of that same end. Such then is the state of that faculty in me which is termed the will; such is the state of those particular acts or modifications of that faculty, which are termed wishes or desires, and which have their immediate efficient causes in corresponding feelings, in corresponding pleasure and pain, such as, on the occasion in question, the imagination brings to view.” Bentham (1827: 4)

state of affairs by introducing a sharp distinction between the information level associated with the individual utilities  $(U_i)_{i \in \mathbb{N}}$  which leads into a formal definition of welfarism, and the outcome level based on  $(V_i)_{i \in \mathbb{N}}$  which leads into the ethical definition.<sup>52</sup> Individual utilities should not be used in advance of any reconstruction from  $U_i$  to  $V_i$  for calculation of the collective social welfare in regard to ethical welfarism as such. As the father of utilitarianism, Bentham is obviously welfarist; but he is not welfarist according to the formal definition (which is the usual attribution within contemporary welfare economics); he is instead welfarist according to the ethical definition.

### **Goodness and Fairness in Bentham's doctrine**

Let us take seriously what is at stake in reconstructing individual utilities before making any ethical judgments. Utilitarianism turns out not to be a theory that can be straightforwardly applied once we have information about individual utilities; first, it is imperative to reconstruct accurate information on these utilities.

Recall that individual utilities  $(U_i)_{i \in \mathbb{N}}$  represent the principle of the Good in any utilitarian theory, among them that of Bentham. The entire utilitarian system is supposed to induce the Good and the Just to merge. If we retain the standard formal definition, this merger is implemented bottom-up—from the individual level to the collective level. Henceforth, Goodness is prior to fairness, as is consistent with the standard conception of utilitarianism.

But  $(V_i)_{i \in \mathbb{N}}$  is and ought to be different from  $(U_i)_{i \in \mathbb{N}}$ .  $(V_i)$  is defined in a top-down way, from the collective level to the individual level. Defining an ethical criterion at the collective level introduces issues of justice rather than of individual morality as such. There is in fact no problem of justice on Robinson Crusoe's island, at least until Friday comes along. There is however clearly a problem with justice at the collective level when one seeks to tackle the problem of distribution among several persons and, above all, when there is also at stake the correction of negative externalities—as for instance in the reformulation of individual utilities into  $V_i$  to take into account the interaction structure of the prisoner's dilemma. As soon as the list  $(V_i)_{i \in \mathbb{N}}$  is taken into account, rather than mere individual interests in themselves without regard to interactions, then justice becomes prior to the good. The priority of fairness over goodness inevitably recalls Rawls's definition of political liberalism.<sup>53</sup> Paradoxically,

---

<sup>52</sup>Notice the proximity of this more nuanced formulation to the distinction between eudemonia and welfarism. See Sen (1979b: 472) or Kolm (2005).

<sup>53</sup>See Rawls (1971, 1985, 1993, 2001).

Rawls's theory of justice as fairness is intentionally opposed to utilitarianism specifically in this regard.

Here we need to distinguish discussion of Bentham's doctrine from that of other utilitarian theories. We shall demonstrate that this irksome paradox is essential for Bentham's doctrine to be consistent, even though others have rather avoided this fact. The merging of individual and collective interests is indeed at the heart of all utilitarianism doctrines. Other contemporary utilitarians did indeed tackle this problem explicitly. Harsanyi (1955, 1992) for instance makes a formal difference between subjective utilities and moral utilities. Those now using the preference model of utility introduced many conditions for relevant preferences in constructing social utility: preferences should be informed, based on true beliefs, and rational.<sup>54</sup> Some even acknowledge the necessity of laundering preferences,<sup>55</sup> rejecting meddling, altruistic, or sadistic preferences, among others. This process of laundering preferences, or the construction of a well-designed ethical utility independent of actual preferences, is a step backwards in respect to strict welfarism. An ontological theory of value is needed to design what would be considered ethical utility, or appropriate laundering.<sup>56</sup> Whenever the reconstruction of individual interest is based on a substantial theory of the Good, as it is in all the above examples, the resulting utilitarian system gives priority to the Good over the Just. We should emphasise the fact that this substantial theory of the Good is exogenous to individual utilities. In other words, these utilitarian theories require not only the two standard elements, welfarism and aggregation (as asserted by Sen (1979b)), but also a 'third element': some idea of what is ontologically good for individuals. When this third element is introduced into the theory, the priority of goodness over fairness emerges without difficulty.

However, this third element is an inevitable step towards natural laws or fundamental rights. Bentham would not accept it. He was explicitly and fiercely opposed to such an idea. One should recall that he described the French Declaration of the Rights of Man as "Nonsense Upon Stilts".<sup>57</sup> But without a third element there is no way to justify exogenously the use of a different notion of individual utility. The required reconstruction of individual interests from  $U_i$  to  $V_i$  is based on a consideration of collective welfare  $W$ —here determined by aggregation—rather than on normative

---

<sup>54</sup>See Haslett (1990) for many relevant references.

<sup>55</sup>See Goodin (1986).

<sup>56</sup>See Nussbaum (2001). Note that the laundering process might suffer from the same criticism regarding paternalist as has been the case with the capability approach.

<sup>57</sup>See his letter to Jacques Pierre Brissot De Warville, mid-August 1789. On this see Cot (1993).

considerations at the individual level.<sup>58</sup> As we claimed above, resorting to  $V_i$  means valuing justice rather than goodness. The conclusion is now straightforward: for Bentham, Fairness is prior to Goodness.

Is Bentham conscious of the priority that his doctrine gives to fairness over goodness, once consistently reformulated? First of all, Bentham does not intend to lend priority to one or the other, his objective is to bring about a merger of both. Hence most of his discussion is devoted to the absence of a gap between the Just and the Good. Bentham wishes to decrease the importance of the State and increase that of the deontologist.<sup>59</sup> The ideal society would be that in which no artificial enforcement were necessary, and where all individuals have incorporated the collective interest as their own. In other words, he wishes that the problem did not exist.<sup>60</sup> That is why the issue, in the absence of externalities, the effect of time, and even for certain configurations of externalities (see above) is not raised. Any conclusion on the subject requires a focus on specific cases where an artificial junction of interests can deal with this problem.

Secondly, there are some hints that Bentham would not have totally rejected such a conclusion. He carefully uses words like “the standard of right and wrong”,<sup>61</sup> “right

---

<sup>58</sup>There is of course a direct link between such consideration and the standard criticisms regarding the sacrifice of minorities and authoritarian aspects of utilitarianism. See Foucault (1975), Baujard (2003: 67–72).

<sup>59</sup>See Bentham (1831a: 37).

<sup>60</sup> “In a good moral condition, the duty of a man should never be to do that which it is not in his interest to do. Morality teaches him to form a proper estimation of his interest and his duties; and in examining he will perceive their coincidence. One is accustomed to say that a man has to do his duty in the face of sacrificing his interests. It is not unusual to hear of this or that individual having made this sacrifice, and there is never any lack of admiration expressed in such cases. But considering interest and duty in a broader perspective one will see that in the ordinary matters of life the sacrifice of interest to duty is not practicable, nor even very desirable; that such a sacrifice is not possible, and that if made it contributes nothing to the well-being of humanity.” *Trans.* “En saine morale, le devoir d’un homme ne saurait jamais consister à faire ce qu’il est de son intérêt de ne pas faire. La morale lui enseignera à établir une juste estimation de ses intérêts et de ses devoirs ; et en les examinant, il apercevra leur coïncidence. On a coutume de dire qu’un homme doit faire à ses devoirs le sacrifice de ses intérêts. Il n’est pas rare d’entendre citer tel ou tel individu pour avoir fait ce sacrifice et on ne manque jamais d’exprimer à son sujet son admiration. Mais en considérant l’intérêt et le devoir dans son acception la plus large, on se convaincra que dans les choses ordinaires de la vie, le sacrifice de l’intérêt au devoir n’est pas praticable, ni même beaucoup à désirer ; que ce sacrifice n’est pas possible, et que, s’il pouvait s’effectuer, il ne contribuerait en rien au bonheur de l’humanité” Bentham (1831a: 17-18)

<sup>61</sup>Bentham (1789: 4)

and proper”, “fairness”<sup>62</sup> where he could, as usual, use “good” or “bad”. Consider in particular this passage (Bentham 1827: 6): “In the eyes of every impartial arbiter, writing in the character of legislator, and having exactly the same regard for the happiness of every member of the community in question, as for that of every other, the greatest happiness of the greatest number of the members of that same community, cannot but be recognized in the character of the right and proper and sole right and proper end of government, or say, object of pursuit. For the designation of the opposite, or reverse of what is right and proper, the term sinister may, in consideration of the relation borne to each other by the two terms, taken in their original physical sense be employed.” And he adds: “*That* being taken for an end, to which it is right and proper that all legislative arrangements be directed, my opinion is, that as far as they go, the proposed arrangements which here follow would be in a higher degree conducive to it than any other could be, that could be proposed in a work which was not particularly adapted to the situation of any one country, to the exclusion of all others.” (Bentham, 1827: 7)). Bentham also insists on the fact that, eventually, the aim of each member is the interest of all,<sup>63</sup> not his own. These doubts support the thesis that Bentham attributes a priority to fairness, when defined as the greatest universal happiness, over individual interests themselves.

In sum, what is at stake at the individual level is not an individual notion of individual interest, defining what is good; it is rather a reconstructed proxy for individual interest that is supposed to be relevant and conducive to the most just situation. The term ‘just’ is first of all imperative in so far as the objective of the society is defined at this level. It is also justified by the fact that the interaction of individuals with diverging interests is taken into account at the initiation of this reconstruction. As a consequence, ‘what is good for the community is therefore good for the individual’. I do not claim Bentham had the intention, explicit or conscious, to give priority to fairness—this would not even be very convincing. But I have shown that such a priority underlies his theory, in the sense that it is required for a consistent version of Bentham’s doctrine to obtain the anticipated merger between the Just and the Good.

#### 4 CONCLUSION

---

<sup>62</sup>Bentham (1789: xxxvii)

<sup>63</sup>“Hence it is, that to serve for all occasions, instead of saying the greatest happiness of all, it becomes necessary to use the expression, the greatest happiness of the greatest number. If however, instead of the word *happiness*, the word *interest* is employed, the phrase *universal interest* may be employed as corresponding indifferently to the interest of the greatest number, or to the interest of all.” Bentham (1827: 6)

I have discussed accepted views of the ‘is’ and the ‘ought’ in Bentham’s works. Utility should guide both, but there are conflicts with the latter. To state this is not to reveal—at least at this point—an inner contradiction in Bentham’s theory; it rather testifies to the complexity of the link between individual and collective interests. I have focused on those situations where a gap exists between individual and collective interest, prompting examination of Bentham’s ‘felicific calculus’. The reconciliation of individual and collective interests requires that a link be established between the positive and the normative levels. As there is not always spontaneously a harmony of interest at the positive level, the utility of objects should be modified to induce people to act in the *right* direction, which direction is defined by the utilitarian ethical criterion. I have put forward a reformulation of the positive and normative analysis of individual and collective interests, and focused on the analytical link between individual behaviours and collective optimality. I then justified the distinction between two notions of individual interest, depending on the level of analysis—individual or collective—or on the ethical view—goodness or fairness. This leads to two main conclusions.

Firstly, we have shown that Bentham’s welfarism is confirmed in the ethical definition, but could not be so confirmed in the formal definition. The relevant information at the individual level cannot be methodologically individualist and has to take into account, for Bentham himself, other alternatives and other people as soon as externalities are taken seriously. The distinction between the  $U_i$  individual interest and the  $V_i$  individual interest clearly questions the common intuition according to which formal welfarism is simply a way of expressing ethical welfarism in mathematics and formal economic argument. As a consequence, we have shown in general that there is no one-to-one relationship between the ethical definition of welfarism and the formal definition.

Secondly, the thesis we have here advanced modifies our perception of Bentham’s theory. “Good” individual interest maximizes individual self-interest; “fair” individual interest maximizes—indirectly—the collective interest when individuals maximize their own. The latter definition is morally superior to the former. In other words, I claim that the father of utilitarianism attributes a priority to fairness over the Good and that fairness is not merely deduced from the Good. This alters the standard view on Bentham’s utilitarianism, and might even disrupt our views about utilitarianism in general. Utilitarians have, as in Harsanyi’s theory, accepted a substantive theory of the Good, such as in the natural right tradition; or, as in Bentham’s theory, commitment to any theory of natural rights is explicitly rejected. In the former case, an external theory of the Good is necessary to create sufficient individual information—as with Harsanyi. In the latter case, a theory of fairness is primarily required—as Bentham. The primary importance of individual interests is therefore called into question, and

the collective interest becomes the foremost issue. Utilitarianism has based its notoriety and its success on the simple idea that it was primarily and essentially a theory which sought to take look after people's welfare, being the only good judge of what is Good. This paper challenging this view, showing instead that utilitarianism is instead based on a substantive theory of the Good or fairness, whether explicit or not. Is this the death knell of utilitarianism? This paper seeks to further debate concerning the possible inconsistency of Bentham's doctrine.

Thirdly, this result contributes to contemporary debate in welfare economics. When defining a public objective, most social welfare functions in the literature are indeed (at least formally) welfarist, meaning they rely (simply) on individual utility functions. A concrete question could be: what kind of utility function should we use for this social welfare function? Should it be the rational consumer's utility function? If the answer to this question is "yes", then a relationship between rational choice theory and normative theory, between the 'is' and the 'ought', would be necessary. It does not mean that 'ought' and 'is' are supposed close or identical, but that the 'ought' is built on a reformulation of the 'is'. What we have shown here is that, because of this ethical link and also a technical relation arising from inevitable interactions between individuals, a further ethical link is built between individual and collective interests: the individual interest relevant for collective purposes is different to that relevant for individual purposes. If we wished to re-interpret Bentham's view *via* a social welfare function, our conclusion would imply the use of utility functions distinct from those of the consumer. This would in turn imply substantial modification of the formal definitions of the utility function or of the preference relation used in the models of contemporary welfare economics.

## REFERENCES

- [1] AUDARD, C. *Anthologie historique et critique de l'utilitarisme*. Presses universitaires de France, Paris, 1999.
- [2] BAUJARD, A. *Bien-être individuel et justice sociale. Fondement, caractérisation formelle et analyse normative*. PhD thesis, Université de Caen Basse-Normandie, 2003.
- [3] BAUJARD, A. A return to Bentham's Felicific Calculus. From moral welfarism to technical non-welfarism. *The European Journal of the History of Economic Thought* 16, 3 (sept. 2009), 431–453.
- [4] BENTHAM, J. *Introduction to the principles of morals and legislation*. Blackwell's political texts. Blackwell, Oxford, 1789. Intr. de W. Harrison, 1967.

- [5] BENTHAM, J. *Traité de législation civile et pénale*, vol. 1 : Principes de législation. Rey et Gravier, Libraires, Paris, 1802. par E. Dumont, 1830.
- [6] BENTHAM, J. *A table of the springs of action*. William Tait, Edimburgh, 1817. in *Bentham's work*, published under the superintendence of his executor John Bowring, Vol.1, 1843(1872).
- [7] BENTHAM, J. *Leading principles of a Constitutional Code, for any state*. William Tait, Edimburgh, 1823. first published in the Pamphleeter, N.44, 1823, ed. in *Bentham's work*, published under the superintendence of his executor John Bowring, Vol.1, 1843(1872).
- [8] BENTHAM, J. *The Constitutional Code*. William Tait, Edimburgh, 1827(1830). in *Bentham's work*, published under the superintendence of his executor John Bowring, Vol.9, 1843(1872).
- [9] BENTHAM, J. *Déontologie ou Science de la morale*, vol. 1. Charpentier, Libraire-éditeur, Paris, 1831. revu, remis en ordre et publié par J. Bowring, traduit par B. Laroche, 1834.
- [10] BENTHAM, J. *Déontologie ou Science de la morale*, vol. 2. Charpentier, Libraire-éditeur, Paris, 1831. revu, remis en ordre et publié par J. Bowring, traduit par B. Laroche, 1834.
- [11] BENTHAM, J. *Bentham's work*, 1872 ed. William Tait, Edimburgh, 1843(1872). 11 volumes.
- [12] BROOME, J. Reply to Kolm. *Social Choice and Welfare* 11, 3 (1994), 199–201.
- [13] BURNE, P. Bentham and the utilitarian principle. *Mind, New Series* 58, 231 (1949), 367–368.
- [14] COT, A. Jérémy Bentham, un “Newton” de la morale. In *Nouvelle histoire de la pensée économique*, A. Béraud and G. Facarello, Eds. La découverte, Paris, 2000, pp. 289–301.
- [15] COT, A. L. Utilitarisme, libéralisme économique et libéralisme politique: Jeremy Bentham et la boîte de Pandore des droits naturels. *Économie et Sociétés, “Histoire de la pensée économique”* 18, 12 (1993), 117–145.
- [16] FOUCAULT, M. *Surveiller et punir*. Tel. Gallimard, Paris, 1975.
- [17] FRIEDMAN, M. *The methodology of positive economics*. The University of Chicago press, Chicago, 1953. pp. 3–46 reproduites in D. M. Hausman, *The philosophy of economics. An anthology*, Cambridge University Press, chap. 10, pp. 210–244, 1984.

- [18] GOLDWORTH, A. Jeremy Bentham: On the measurement of subjective states. *The Bentham Newsletter*, 2 (mars 1979), 2–17. The Bentham Committee University College, London.
- [19] GOODIN, R. E. Laundering preferences. In *Foundations of social choice theory*, J. Elster and A. Hylland, Eds. Cambridge University Press, Cambridge, 1986, ch. 3, pp. 75–101.
- [20] GUIDI, M. E. L. Review of ‘Nathalie Sigot, Bentham et l’économie. Une histoire d’utilité.’ *Utilitas* 16, 1 (march 2004).
- [21] GUIDI, M. E. L. On some asymmetries between pain and pleasure: Bentham ‘felicific calculus’ and his critique of Locke and his followers. In *Handbook of Happiness*, L. Bruni and P. Porta, Eds. E. Elgar, Cheltenham, 2007, ch. 4.
- [22] HALEVY, E. *La formation du radicalisme philosophique*, vol. I : La jeunesse de Bentham, 1776-1789. Presses Universitaires de France, Paris, 1901.
- [23] HALEVY, E. *La formation du radicalisme philosophique*, vol. II : L’évolution de la doctrine utilitaire de 1789 à 1815. Presses Universitaires de France, Paris, 1901.
- [24] HALEVY, E. *La formation du radicalisme philosophique*, vol. III : Le radicalisme philosophique. Presses Universitaires de France, Paris, 1904.
- [25] HARRISON, R. *Bentham*. The arguments of the philosophers. Routledge & Kegan Paul, London, Boston, Melbourne and Henley, 1983.
- [26] HARRISON, W. ‘Jeremy Bentham: Ten critical essays’. edited by Bikhu Parekh. (London: Frank Cass, and Portland Oregon: International scholarly book services, 1974. *The American Political Science Review* 71, 2 (1977), 654–655.
- [27] HARSANYI, J. C. Cardinal welfare, individual ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63, 4 (1955), 309–321.
- [28] HARSANYI, J. C. Game and decision theoretic models in ethics. In *Handbook of game Theory*, R. J. Aumann and S. Hart, Eds., vol. 1. Elsevier Science Publisher, 1992, ch. 19, pp. 669–707.
- [29] HASLETT, D. What is utility? *Economics and Philosophy* 6 (1990), 65–94.
- [30] HUME, L. J. Myrdal on Jeremy Bentham: Laissez-faire and harmony of interests. *Economica* 36, 143 (august 1969), 295–303.

- [31] KEETON, G. W., AND SCHWARZENBERGER, G. Jeremy Bentham and the law. A symposium edited on behalf of the faculty of laws of University College, London. Stevens & Sons Ltd., London, 1948.
- [32] KOLM, S.-C. Macro-justice. The political Economy of Fairness. Cambridge, 2005.
- [33] LAVAL, C. Le “sense of interest” de Jeremy Bentham. *Economie et Sociétés* 18, 12 (1993), 147–169.
- [34] MITCHELL, W. C. Bentham’s felicific calculus. *Political Science Quarterly* 33, 2 (1918), 161–183.
- [35] MONGIN, P. L’utilitarisme originel et le développement de la théorie économique. In *La Formation du Radicalisme Philosophique*, Elie Halévy ed., vol. 3 : Le Radicalisme. Presses Universitaires de France, Paris, 1995, ch. Postface, pp. 369–394. Également paru in *La Pensée Politique*, tome 3, pp. 341–361, Gallimard et Le Seuil, Paris, 1995.
- [36] MONGIN, P., AND SIGOT, N. Halévy’s Bentham is Bentham. *Philosophy* 74 (1999), 271–281.
- [37] MUNDLE. ‘Jeremy Bentham and the law.’ A symposium edited on behalf of the faculty of laws of University College, London by George W. Keeton and George Schwarzenberger. London: Stevens & sons Ltd., 1948. *Mind* 58, 229 (1949), 101–130.
- [38] MYRDAL, G. *The political element in the development of economic theory*. Transaction Publishers, 1990, New Brunswick and London, 1932. translated by Paul Streeten from the German edition.
- [39] NUSSBAUM, M. C. Adaptive preferences and women’s options. *Economics and Philosophy* 17 (2001), 67–88.
- [40] RAWLS, J. La théorie de la justice comme équité : Une théorie politique et non pas métaphysique. *Philosophy and Public Affairs* 14, 3. Trad. fr. par C. Audard, *Individu et justice sociale*, Point, Seuil, Paris, pp. 279–317, 1988.
- [41] RAWLS, J. *A Theory of Justice*. The Belknap Press of Harvard university Press, 1971. Trad. fr. par C. Audard, *Théorie de la justice*, Points Seuil, 1997.
- [42] RAWLS, J. *Political Liberalism*. Columbia University Press, 1993. Trad. fr. de C. Audard. *Libéralisme politique*, Presses Universitaires de France, 1995.
- [43] RAWLS, J. *Justice as fairness. A restatement*. Belknap, Harvard, 2001.

- [44] SEN, A. K. Personal utilities and public judgements : Or what's wrong with welfare economics? *The Economic Journal* 89, 355 (septembre 1979), 537–558.
- [45] SEN, A. K. Utilitarianism and welfarism. *The Journal of Philosophy* 76, 9 (1979), 463–489.
- [46] SEN, A. K. Minimal liberty. *Economica* 59, 234 (may 1992), 139–159.
- [47] SHACKLETON, R. The greatest happiness of the greatest number: the history of bentham's phrase. *Studies on Voltaire and the eighteenth century* 90 (1972), 1461–1482. Also published in *Bentham: Moral Political and legal philosophy*, Vol. I : Moral and political philosophy, ed. by Gerard J. Postema. The international library of critical essays in the history of philosophy. Ashgate, Dartmouth, 2002. pp. 29-50.
- [48] SIDGWICK, H. *The methods of ethics*. Macmillan, 1874.
- [49] SIGOT, N. 'Be quiet', mais modérément : Le rôle de l'Etat dans la pensée économique de Jeremy Bentham. *Revue Économique* 44, 1 (janv. 1993), 23–50.
- [50] SIGOT, N. Bentham et l'économie. Une histoire d'utilité. *Economica*, Paris, 2001.
- [51] SPRIGGE, T. L. S. The relation between Jeremy Bentham's psychological and his ethical hedonism. *Utilitas* 11, 3 (nov. 1999), 296–319.
- [52] VERGARA, F. A critique of Elie Halévy. *Philosophy* 73, 1 (1998).
- [53] WARKE, T. Classical utilitarianism and the methodology of determinate choice, in economics and in ethics. *Journal of Economic Methodology* 7, 3 (2000), 373–394.
- [54] WARKE, T. Multi-dimensional utility and the index number problem: Jeremy Bentham, J.S. Mill and qualitative hedonism. *Utilitas* 12 (2000), 176–203.