



HAL
open science

Utilitarianism and anti-utilitarianism

Antoinette Baujard

► **To cite this version:**

| Antoinette Baujard. Utilitarianism and anti-utilitarianism. 2013. halshs-00906899

HAL Id: halshs-00906899

<https://shs.hal.science/halshs-00906899>

Preprint submitted on 20 Nov 2013

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.

WP 1332

Utilitarianism and anti-utilitarianism

Antoinette Baujard

November 2013

GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne

93, chemin des Mouilles 69130 Ecully – France

Tel. +33 (0)4 72 86 60 60

Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France

Tel. +33 (0)4 77 42 19 60

Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : gate@gate.cnrs.fr

Téléchargement / Download : <http://www.gate.cnrs.fr> – Publications / Working Papers

Utilitarianism and anti-utilitarianism*

Antoinette Baujard †

November 8, 2013

Abstract

This paper presents the different utilitarian approaches to ethics. It stresses the influence of utilitarianism in economics in general and in welfare economics in particular. The key idea of the paper to explain the evolution from classical utilitarianism to preferences utilitarianism and towards post-welfarist approaches is the following. Utility is defined normatively and positively. This generates some serious tensions. Utilitarianism needs to evolve to go beyond this ethical tension. Another idea defended in this paper is that the solutions developed by utilitarianism to solve the ethical issue eventually reinforces operational problems. This raises a legitimacy issue as whether the intervention of utilitarian economists in public decision are likely to be normatively transparent.

JEL Codes: A13,A33,B10,B2, D63

Key-words: Utilitarianism, Welfarism, Hedonism, Preference, Utility

Introduction

While there is wide variation in utilitarian approaches to ethics, they are united by their endorsement of the following general principle: the morally right action is judged through the goodness of its outcomes

*To be published in Gilbert Faccarello and Heinz Kurz Eds. *Handbook of the History of Economic Analysis*, Edward Elgar Publishing Limited, Cheltenham (U.K.), Vol.3: Developments in major fields of economics

†GATE L-SE, Université de Lyon, Lyon, F-69007 France; CNRS, GATE Lyon Saint-Etienne, Ecully, F-69130, France; and Université Jean Monnet, Saint-Etienne, F-42000, France. Contact: antoinette.baujard@univ-st-etienne.fr, (33) +4 77 42 13 61.

for society, and, conversely, what is good for society is based on what is good for individuals.

Utilitarianism, as a family of philosophical theories, has been the most powerful and pervasive approach in the development of economics since the marginalist revolution. Utilitarianism was developed in the 18th century and then fully articulated in the 19th, designed to do good to the world not from the point of view of the Christian church or any religion, but rather from a secular point of view based on rational thought. It was conceived as a way to think about the legal system, and to improve it, on the basis of a single coherent rational and acceptable principle, that of the utility principle. It hence may be used to think about the constitution, about civil and penal laws, and, last but not least, any kind of policy judgment whether economic or social. Economics has endorsed some important aspects of utilitarianism ever since the 18th century. In particular, welfare economics, and hence virtually every public policy recommendation formulated by economists, has for years been influenced by utilitarianism in some manner, albeit not always explicitly recognized. The efforts to get away from utilitarianism may even explain aspects of the evolution of welfare economics (See Baujard 2012).

Utilitarianism is not only a moral and political philosophy, however; it is also a philosophy of action. As an ethical theory, it sets down what individuals should do to improve their own situation, and what should be done by every individual and by the collectivity to improve collective welfare. As a theory of action, utilitarianism claims that individuals seek to promote their own utility, such that utility appears as both an explanation of and a guide for human action. Notice that utilitarianism is sometimes claimed to be rooted in a consideration of strictly selfish actions, with no attention paid to the utility of others: but this is merely an exaggerated caricature of the theory. Being motivated by one's own pleasure does not require that this pleasure be exclusively self-oriented. There certainly exist extra-regarding pleasures, functioning as truly motivating factors, such that seeking one's own self-interest does not imply that one disregard the fate of others. The ideational evolution in the utilitarian philosophy of action is exactly parallel to that in utilitarian moral philosophy. In the classical versions of utilitarianism, pleasure guides human actions, as a psychological hedonistic law would suggest. The theory has then moved away from hedonism to consider that the psychological law at stake is the search for the satisfaction of preferences, no matter what

these preferences are made of, and no matter what their substantive cause. We have resolved not to discuss other aspects of the utilitarianist philosophy of action in this paper, in so far as this is closely connected to standard rational choice theory as used in economics: we refer the interested reader to decision theory and the associated heterodox literature.

As a way of setting out the diversity of utilitarian approaches, this chapter has chosen the following as a key idea. Utility has a plural nature: it is both positive and normative, and it is normative not only for every individual, but normative also for the collectivity. This plurality generates tensions. The effort to reconcile these tensions explains the diversity of utilitarian doctrines as well as their evolution, both within classical utilitarian theories and contemporary theories, and even the development from hedonist utilitarianism to preference utilitarianism.

Classical utilitarianism mostly retained a hedonistic interpretation of utility. It is a doctrine that, in its standard 19th century formulation, meant the promotion of the greatest happiness for the greatest number (Section 1). Contemporary utilitarianism can be defined as the combination of act consequentialism, welfarism, and a principle of sum-ranking (Sen 1979c). Consequentialism implies that an action is moral if and only if the social outcome of the resulting state of the world is good. Welfarism is the principle that the goodness of an outcome depends solely on individual utilities and on no other information. Sum-ranking says that the appropriate method of aggregation is to add individual utilities. Contemporary approaches in utilitarianism develop refinements of preferences utilitarianism (Section 2). There are also a wide range of anti-utilitarian theories, which do not in general question the premiss that more goodness is better than less, but rather question the ethical implications of monism—the exclusive focus on utility as opposed to other values—the priority of goodness over fairness, or the democratic failures of utilitarianism. In the economics literature, the alternative to utilitarianism is often supposed to be represented by the Rawlsian theory, as generally encapsulated in the famous Bentham–Rawls opposition. Rawls has indeed been important in making it acceptable to call into question the assertion of an all-powerful and uncontested utilitarianism. There now exists a diversity of critiques and alternatives to utilitarianism (Section 3).

1 Early and classical utilitarianism

The proto-history of utilitarianism goes back to the British moralists of the 17th century, including Richard Cumberland and John Gay, who defended the link between individual morality and the general good. The early utilitarians were influenced by the materialist Helvétius, who claimed that interest dictates judgments, and stressed the importance of education in the formation of human souls; as well as by Francis Hutcheson and David Hume, who held that a kind of commonsense morality is generally compatible with public utility. (For more details on the precursors of utilitarianism, see *e.g.*, Rosen 2003, Driver 2009).

For further discussion of these authors, we refer the readers to the relevant chapters in the handbook edited by Faccarello and Kurz. Our aim in this section is rather to present the historical stages of classical utilitarianism, defined as follows:

The utilitarian doctrine is, that happiness is desirable, and the only thing desirable, as an end; all other thing being only desirable as means to that end. (Mill 1861: 15)

1.1 Utilitarian hedonism

Bentham (1748–1832) is generally considered the father of actual utilitarianism. The principle of utility—the “greatest happiness principle” or “the greatest happiness for the greatest number”—is the fundamental utilitarian principle, as expressed in this famous passage:

Nature has placed mankind under the governance of two sovereign masters, pains and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all we do, in all we say, in all we think [...] The principle of utility recognizes this subjection, and assumes it for the foundation of that system, the object of which is to rear the fabric of felicity by the hands or reason and of law. (Bentham 1789: 1–2)

Bentham sometimes alludes to the sum of individual utility, but the commitment of classical utilitarianism to the principle of the sum does not seem as robust and important as that of two other essential prin-

ciples: impartiality and welfarism. Impartiality means that everyone counts for one and none for more than one. Welfarism is the focus on individual utility. Utility is conceived as a mental state representing happiness, the positive result of a balance between pains and pleasures or any personal valuable feelings. Individual utilities are calculated in three stages (Baujard 2009). First, Bentham lists the different classes of elementary pleasures or pains. Second, the value of utility is assessed according to different criteria including intensity, duration, certainty or uncertainty, propinquity or remoteness, fecundity, purity and extent. Finally, the result of this calculation is adjusted to each individual on the basis of the circumstances that influence her sensitivity. Since the idea of measuring pleasures and comparing them raises many practical problems, Bentham proposes to use some proxy for estimating it, and money may be able to provide the needed measuring rod—this being the ancestor of the idea of the willingness to pay or to accept. More generally, there is a global assumption according to which everyone shares vital concerns including security of expectations, subsistence, abundance and equality, such that rights are associated to all these goals.

A tension lies in the duality of utility, which is supposed both to explain and justify actions. Why would an action that is desired by an individual always be good for her? And, moreover, why would it be good for society? Is vice really undesirable and virtue desirable? Should individuals sacrifice their happiness for that of others? There can be tensions between the reach of the individual's happiness and that of society, which utilitarianism shall need to reconcile.

Bentham recognizes there may be mistakes, cognitive limits, or more fundamentally what we would call now external effects—as introduced by the criteria of fecundity, purity and extent— that are likely to hamper the ideal conjunction of individual interests. Whenever utility is likely not to be maximized, Bentham designs external punishments and rewards, and appeals to various kinds of influence (such as peer pressure) to make sure that individuals are incited to revert to a better action, so that egoistic people, or people who might be wrong, bad at calculating, or insufficiently informed, will eventually act to maximize social utility.

John Stuart Mill's utilitarianism (1806-1873), influenced by perfectionist intuitions, is also meant to answer objections such as those levelled at Bentham's theory. Among other insights, he recognizes that all pleasures do not seem equally valuable, as evoked in a celebrated

passage:

It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied. And if the fool, or the pig, are of a different opinion, it is because they only know their own side of the question. (Mill 1861: 37)

Mill's utilitarianism is able to discriminate higher moral and aesthetic sentiments from sources of swinish pleasure. His proof for his principle of ranking is based on the fact that those who have experienced both kinds of pleasures know which is more valuable; this is why we speak of Mill's qualitative utilitarianism.

The details of the consideration of different sources of pleasure as set out in Mill's view has led certain authors to regret Mill's "naturalist fallacy" (Moore 1903). In order to distinguish between different kinds of pleasures, we need to suppose there exists another value, other than utility, which has intrinsic importance, such that some specific actions are good by their nature rather than by their actual consequences.

1.2 From utilitarianism to economics

Henry Sidgwick's (1838–1900; See Vol. 1) *The methods of ethics* has been influential in the clarification of important features of utilitarianism and as regards some developments of the marginalist literature in economics. First, he shows that utilitarianism can resolve any conflicts of values or rules, at least under the principle that an ethical agent must be impartial between one person's pleasures and another's. Sidgwick, though, made an important case for criticizing the hedonist approach in utilitarianism: a problem of "dualism of practical reason" occurs when it seems better on utilitarian grounds to sacrifice one's own interests for others'. Some economists, notably Jevons (1871), and more recently Skyrms (1996) among many others, have answered this problem by defending the claim that utility was made a moral norm through socio-evolutionary explanations of the emergence and reinforcement of utilitarian rules, so that the existence of a link between the normative and the positive approaches to utility does not seem implausible.

Second, Sidgwick discusses the impact of time and population on utilitarianism. He tackles the problem of the sacrifice of a present pleasure for a greater one in the future, seeing this as parallel to the

sacrifice of one's own pleasure for somebody else's. Francis Ysidro Edgeworth (1845–1926) assumed that natural units of pleasure and pain can in principle be ascertained and aggregated over varying populations and time horizons. The population issue later came to be the subject of extensively study (Blackorby, Bossert and Donaldson 2005). In modern utilitarianism, social welfare is optimized when the sum of individual cardinal utilities is maximized. Derek Parfit (1984) showed that sum utilitarianism was exaggeratedly populationist, because it implies what he called 'the repugnant conclusion': that a numerous and miserable population is to be judged better off than a sparse and happy one. Conversely, average utilitarianism would be Malthusian because new inhabitants provide a distinct value according to whether they are more or less happy.

Third, Sidgwick stressed the need for developing formal models of a utilitarian calculus under ideal conditions so that the implications of quantitative hedonistic utilitarian reasoning could be clarified. The utilitarian economists have gone beyond his original intentions. They have assumed that the utility numbers are known with precision, which allows them to compute properly interpersonal comparisons of utilities and the sum of utilities in every situation. An issue is thus raised as to who shall be responsible for performing and enforcing the utilitarian calculations. Edgeworth and others have considered it best to trust a utilitarian elite for this delicate task. Along the same line, Sidgwick made a point of insisting that the utilitarian theory should not be publicized, because uneducated people, those incapable of moral reflection, may misuse it. Utilitarian rules should thus rather be applied by educated people, while the laymen should be content to be governed by the utilitarian elite. This corresponds to the colonial elitism of Sidgwick's times, which Bernard Williams (1973) derisively called "Government House Utilitarianism". It is hard to deny that utilitarianism in this sense utterly contradicts individual sovereignty, at least that of allegedly uneducated persons.

2 Contemporary utilitarianism

The ordinalist revolution that affected economic science from Pareto to the thirties watershed has totally modified the definition of utility in economics: in order to secure economics' status as a science, it was necessary to exclude the hedonist interpretation of utility and favor

the preference interpretation; this move was completed by the revealed preference theory (Baujard 2013). A similar evolution occurred in the utilitarian philosophical literature, especially in the second part of the twentieth century. Both in economics and in philosophy, the model of utility as a measure of pleasure has now been replaced by the preference model (on the comparison of the two models, see *e.g.*, Haslett 1990). Preference utilitarianism also corresponds to a preference view of welfare: if you prefer x to y , it also follows you are better off in x than in y .

2.1 Harsanyi's rational choice utilitarianism

Harsanyi's (1953,1955) articles are both leading contributions to modern utilitarianism. The 1953 paper offers a philosophical reconstruction of the notion of utilitarian impartiality. Harsanyi imagines that an impartial spectator compares the utility of two populations. The fiction ought to help derive a function of social welfare from individual utilities, capturing the fact that social utility is based on the individuals' personal assessment of their situation. A primary requirement is that the spectator should be impartial, that is to say, he should not favor the situation of one individual over another, whatever their differences. For instance, a poorer person should not be favored over a richer one, nor the converse. Everyone is hence given a strictly equal weight. Harsanyi's innovative idea was to extract this reasoning from the context of risk theory. Say you were to become one person in a given population: you have as much chance of becoming one person as any other—as much chance of being rich as poor. From outside, and not knowing who you end up eventually to be, you have no reason to favor one person over another in revealing your preference, so that you should consider being any person in the society with equal chance. You therefore want to maximize the sum of all individual utilities—for a given population, equivalently, the average utility. Impartiality is enforced by the fiction of a (thin) veil of ignorance.

In his 1955 article, Harsanyi provides an axiomatic justification of the criterion of the sum. Based on the von Neumann and Morgenstern characterization of utility under risk, Harsanyi treats the individual reasoning under the veil of ignorance we described above. For any weights a_i , and for utility u_i of all the individuals i of the society composed of N individuals, he obtains the following characterization

of weighted utilitarianism:

$$W = \sum_{i \in N} a_i \cdot u_i$$

This “construction” of what utility should be under impartiality entails there exist two definitions of utility. Actual utility—or subjective utility—corresponds to a numerical indicator of preferences, which is revealed by actual choice behavior. Such utilities are ordinal, and neither cardinal measures nor comparisons of utility in this sense are meaningful. By contrast, moral utilities are cardinal and interpersonally comparable. The identification of moral utility requires certain conditions to enforce impartiality. This is the task of the veil of ignorance. Furthermore, imagine that i is an altruistic person, whose utility is increasing with j 's utility. Then j 's utility is likely to be counted twice, in contravention of the impartiality criterion. Other-regarding preferences should therefore be excluded in order to avoid such double-counting (Dworkin 1978).

Since their publication, however, the implications of Harsanyi's results have spilled over into diverse issues: the interpretation of possible negative weights, problems of distribution, cardinality, difficulty with the introduction of subjective and heterogeneous utility, inconsistencies, etc. Here we restrict ourselves to alluding to one of the most famous outcomes, the rich debate between Rawls and Harsanyi (*e.g.*, Rawls 1974, Harsanyi 1975). The reasoning under uncertainty on which Harsanyi's utilitarian social welfare function is based relies heavily on the specifics of his choices regarding risk aversion. Imagine you do not even know the list of all possible situations: you then face a situation of uncertainty rather than a risky situation. You can hardly weight all the possible situations equally, while you may fear falling into the worst situation. John Rawls (1971) considered that the latter fiction was a better fit for his criterion of impartiality. He proposed a thick veil of ignorance, such that the maximin criterion would be more suitable. In the end, utilitarian impartiality implies the maximization of the sum of utilities, while the Rawlsian theory of justice as fairness entails devoting all attention to the most disfavored group in society. As a consequence, the former ignores all distributive issues while the latter is fundamentally egalitarian.

2.2 Preference utilitarianism

Most modern contributions to utilitarianism retain the preference model of utility, such that utility is simply the satisfaction of preferences. The ‘strong preference’ model is the one that is most commonly used in social sciences: individual preferences are revealed in the actual choices of individuals. Specifically, utilities are the numerical representation of preferences as revealed by choices. If I choose wine when I could drink tea, my preference is for wine over tea. In terms of utility, the numerical value of the utility that I associate with wine is therefore greater than the one I associate with tea. For example, I can capture either the choice situation or the preference by the following utility numbers: $U(\text{Wine}) = 10$ and $U(\text{Tea}) = 0$ or, equivalently, 4 or -2.67. In this context, ordinal measures of utility are all that counts, but cardinal utility is meaningless—such that summing and comparison are impossible. Arrow and Hahn, Debreu, Hicks, Samuelson, among many others, had been convinced of the primacy of this model for economic science since the forties. The revealed preference model is the basis for the interpretation of the utilities found in our microeconomic textbooks. Nevertheless, the same model of utility is likely to be used in other chapters of these textbooks, those which tackle issues of welfare economics, and even applied public economics. In this context, social welfare is hence supposed to depend on individual utilities alone, which only represent individual choices. The use of this model in welfare economics is heavily criticized for several reasons (*e.g.*, Hausman 2012). For one thing, it ignores the fact that individuals may act against their own interests: for instance, a knowledgeable person who continues to smoke, or a penniless young man who prefers to work for a low wage rather than concentrate on his studies. More generally, we cannot distinguish actions by interest, lack of information, whether they are based on false beliefs or lack of will, constrained by social norms, etc. As a consequence, it is difficult to interpret the concept of utility corresponding to the strong preference model as the well-being of the person, and even more as relevant information to compute social welfare.

While the strong model of preferences—*i.e.*, revealed preference theory—is mostly retained in economics, most philosophers prefer to retain a weak model of preferences, as we shall now set out. They claim that utility stands for specific kinds of preferences: they should be rational, informed, and based on true beliefs (*e.g.*, Hare 1981, Grif-

fin 1986, Sumner 1996). The introduction of these conditions has two consequences. First, this model answers most of the criticisms that plagued utilitarianism both in its hedonist version and in the strong preference model. Second, it implies that, for utilitarianism, the relevant preferences are not actual preferences: the latter should rather be ‘laundered’ (Goodin 1986)—notice that the issue of who should be entitled to perform the actual task of laundering preferences when utilitarian policies are implemented remains open so far.

Further, if you suppose we know the rational and informed preferences, and if you can transform each individual utility by any positive monotonic function $f(\cdot)$, then you can design some ordinalist utilitarianism where:

$$W = \sum_{i \in N} f(u_i(x))$$

It has been shown that the Rawlsian ranking is a special case of all utilitarian ordinalist functions, where the concavity of f is extreme (Arrow 1973).

3 Anti-utilitarianism

The scope of criticism of utilitarianism is wide, whether on ethical grounds or on practical issues. As we shall see, such criticism can be accommodated by utilitarianism with refinements. However, this critical movement has also induced a sort of ‘non-utility revolution’, both in political philosophy and normative economics. Notice that most alternative approaches would not call themselves anti-utilitarian—though a notable exception is the M.A.U.S.S. project (in French, the “Mouvement Anti-Utilitariste dans les Sciences Sociales”; see, *e.g.*, Caillé 2006)—but at most post-welfarist.

3.1 Ethical limits of utilitarianism

The principle of aggregation has been criticized for being a mere sum of utilities which rules out any distributive considerations. Indeed, in a crude version of utilitarianism, it is considered socially equivalent to give a large amount of money either to one rich person or to a myriad of poor people who would be able to change their quality of life thanks to it. This example runs against moral intuitions and should induce the rejection of the theory. Yet it seems implausible to claim that

equality has never been a concern of utilitarian thinkers. In a old and standard assumption, that goes back not only to the precursors of the marginalists, but also to the first utilitarians, the marginal utility of money is decreasing. In order to maximize the sum of individual utility, it is therefore socially better to give an extra euro to a poor person than to give the same extra euro to a rich one: for the former act is a bigger producer of utility than the latter. A way of representing the egalitarian functions of social welfare is specifically to retain a utilitarian utility function with concave individual utility functions. In this sense, the criticism over the issue of distribution holds only for crude versions of utilitarianism where one euro is supposed to provide the same amount of utility to any person; yet this case is more important than we might think at first sight, in view of Robbins's claim that the assumption of decreasing marginal utility was a normative one and should consequently be kept away from economic science. Nevertheless, even with some given concave utility function, the choice of the criterion of sum (or average) may have consequences that collide with distributive concerns. Utilitarianism can judge certain inequalities as socially better in so far as they contribute to increasing the sum of utilities. For example, it may be appropriate to give subsidies to a rich businessman rather than using the same amount of money to improve the fate of worse-off children if the extra utility associated with the grant is higher than that associated with assistance. Some utilitarians are comfortable with this criticism since it implies, after all, an increase in the sum of utilities: the decisions of the businessman, they say, may provide more jobs and wealth in a wide area, increasing the welfare of numerous families in a way likely to compensate for the low quality of the children's lives. Opponents of utilitarianism then return with another example, where the unquestioned new increase of welfare only accrues to families that were already well-off. While the choice of examples in this vigorous debate seems crucial to determine which side is more convincing, it remains that there exist cases where the utilitarian criteria shall favor the rich over the poor, the well-off to the needy. In the end, it remains that distributive issues may be important in utilitarianism, yet only on instrumental bases. This dependency on utility as the sole intrinsic value still generates unquestionable violations of equality concerns.

Beyond the issue of equality, the utility principle is likely to justify the sacrifice of minorities. In a famous example, utilitarianism underwrites the Roman spectacle of the circus. In the arena, a few Christian

martyrs suffer a great deal, while each spectator derives a little pleasure from observing their suffering. If the spectators are sufficiently numerous, the sum of their low positive utilities shall compensate for the pain of the few Christians.

More generally, the focus on utility implies possible violations of any liberal rights. According to an interpretation of Sen's theorem of the Paretian liberal (1970), it is impossible to attribute an intrinsic importance to utility while simultaneously endorsing a specific concept of freedom. This interpretation has been generalized on the basis of other results pertaining to utility and other values.

Sen has called "welfarism" the doctrine according to which social welfare only depends on utility and on no other values or information (Sen 1979a,b). All versions of utilitarianism are fundamentally welfarist. But welfarism is a problem firstly because it implies the rejection of any other values (or at least it subsumes them). Some alternative theories consider that a plurality of values should be at stake for welfare issues: they make a case for a multidimensional account of welfare, such as quality of life or capabilities (Nussbaum and Sen 1993), complex equality (Walzer 1983), etc. Secondly, utility may not be the suitable information for gauging social welfare, because it entails counterintuitive consequences. After Sen's famous paper "Equality of what?" (Sen 1979c), it has been common to consider alternative informational bases of justice such as primary goods (Rawls 1971), resources (Dworkin 1981a,b), access to advantages (Cohen 1989), welfare opportunities (Arneson 1989), opportunities (Roemer 1999), etc.

Utilitarianism considers that utility, which is important for individuals, is also the relevant information for resolving issues of justice. Some regret the confusion of justice and morality in utilitarianism, where goodness is prior to fairness. Conversely, in other traditions—mainly political liberalism—morality focuses on every individual's judgement of what is good or bad; justice concerns all society and tackles the issue of living together, no matter what the individuals' differences, and even considering divergences in their moral views. Consider the usual case of conflict of interests: utility, even if perfect for one individual, is hardly suitable for improving social welfare. The redefinition of welfare imposed in utilitarianism in order to circumvent the absence of natural conjunction of interests raises fundamental problems regarding the integrity and the autonomy of persons. Firstly, utilitarianism may force individuals to sacrifice something to benefit others, including total strangers. For instance, utilitarianism

is likely to justify the execution of innocent people to prevent rioting and consequent numerous deaths, where raising the general welfare is presented as a serious justification for the innocent's deaths. Secondly, utilitarianism imposes negative responsibilities. Individuals are as responsible for what they do not do as for what they do: people have the responsibility to pursue better social consequences, and not just better consequences for themselves (Williams 1973). It is debatable, however, on what basis utilitarianism can impose upon people a goal that is not a priori supposed to be theirs. Thirdly, Rawls (1971: 26–27) claims that utilitarian neutrality implies the loss of the essential separateness of individuals, such that some human beings may be used as a means to the welfare of others.

On [a utilitarian] conception of society separate individuals are thought of as so many different lines along which rights and duties are to be assigned and scarce means of satisfaction allocated in accordance with rules so as to give the greatest fulfillment of wants. [...] The correct decision is essentially a question of efficient administration. [...] Utilitarianism does not take seriously the distinction between persons. (Rawls 1971: 27)

This argument quickly became famous within the debates over utilitarianism, and has been greatly discussed and refined (see, *e.g.*, Parfit 1984). Rawlsian political liberalism, in which fairness is prior to goodness, constitutes the most famous alternative to utilitarianism.

Fourthly, another family of ethical criticisms concerns the democratic failures of utilitarianism when its implementation is at stake, as we shall have cause to see below.

3.2 Practical limits of utilitarianism

Though welfarism, and especially welfarism based on subjective utility, is widely used in normative economics, the empirical facts hardly confirm that people are indeed utilitarian. Nozick (1974) imagined an “experience machine”, which has been fatal to hedonist utilitarianism. Individuals are offered the chance to be plugged into a machine for life. The latter would feed signals into their brain so that they felt wonderful experiences, irrespective of what was actually happening in the world. Since people do not (in general) prefer the happiness machine to their actual life, it seems to follow that the mental state

of happiness is not everything that is valued by these people. The preference model does not do better in this regard. In their seminal article on normative experimental economics, based on a survey, Yaari and Bar-Hillel (1984) highlighted in particular that the assessment of income distributions turned on needs rather than on tastes or beliefs, as utilitarians would claim. According to an important and robust result of normative experimental economics, the way people judge the fairness of situations depends on contextual circumstances and not on the mere description of individual utilities. For example, for similar utility levels, an even split of a resource between two persons is not judged as being equal when survey respondents learn that one recipient is hard working and disabled while the other is lazy and already in receipt of benefits. Utility thus does not seem in fact to be the sole important value, and nor is it the sole relevant source of information in people's eyes.

Notably defended by S. E. Toulmin and J. Austin, rule utilitarianism had already been clearly formulated by J. O. Urmson and R. B. Brandt in the fifties. It supposes that the consequences of rules are assessed, and that actions—and also therefore classes of actions—are judged, according to rules. Harrod (1936) raised an important problem with rule utilitarianism on the basis of a famous example. A lie is a bad thing, and a good rule should be not to lie. But in special cases a lie is likely to induce better outcomes—although it would be judged negatively by rule-utilitarianism, contrary to a priori intuition. Accordingly, act utilitarianism, where every action is assessed through its actual consequences, is favored. This other version of utilitarianism was defended originally by Bentham, Sidgwick and Moore, and by J. J. C. Smart in the debates of the fifties. Against act-utilitarianism, however, some have complained that it is unable to guide practical decision making: because of lack of information, because gathering such information would be too tedious and costly, because mistakes are more likely to be made when there are so many calculations to be done, and because existing rules or norms could not be used to build expectations nor in particular to trust others' promises. In a nutshell, rule-utilitarianism requires approximations and generalizations which are likely to conflict with the utility principle in certain concrete cases, while act-utilitarianism requires unworkable computations.

Notice that the possibility of measuring utility and implementing utilitarian policies depends on the choice of the utility model. If utility were a mental state, it would be quite difficult to consider the different

natures of pleasures on a common scale, in so far as many a priori seem incommensurable. The model of experience in play here also entails some counter-intuitive consequences. For example, as long as Mary does not know that she has been ruined on Black Tuesday of the 1929 Crash, she has not experienced her misfortune. Whereas it is impossible to deny that her utility—as experienced happiness—remains as high as when she was a wealthy lady, it is also hard to accept that she remains as well off as before. Moreover, it is impossible—or, rather, difficult—to measure, compare, and add feelings, *i.e.*, subjective utilities. As we have already underlined, the preference model supposes individual ordinal rankings that are not easier to compare nor to add. This again hampers actual implementation of utilitarian policies.

A famous criticism of utilitarianism is based on Michel Foucault (1975)'s analysis of Bentham's Panopticon. The Panopticon is a "simple idea of architecture" (Bentham 1791: 5), which determines the organization of life in a prison as a first step. Bentham's project was to apply this in many other areas of social life, such as homes for the poor, schools, hospitals, public administration, factories, etc. The prison cells are arranged concentrically around a central building where an inspector is located: he personifies the otherwise fictitious impartial spectator. He can observe each person's action, and everything that happens. Prisoners are kept separate from the others and from the spectator, of whom they know only they might be being observed.

Hence the major effect of the Panopticon: to induce in the inmate state a conscious and permanent visibility that assures the automatic functioning of power. (Foucault 1975: 234)

The inspector reflects the behavioral expectations of the whole society, and the prisoners feel in their flesh the pressure of the utility principle: hence their own autonomous desires are transformed now so as to feel and act in order to improve the greatest happiness for the greatest number: individuals gradually become their own jailer. As all that is needed here is the thought that there might be an inspector, who may be anybody: he may be changeable or even absent, such that the system seems free from any drift towards tyranny—precisely in accordance with the requirement of impartiality. It remains that individual autonomy is here being totally neglected in favour of social utility: the disciplinary system organized by the Panopticon guarantees the total

submission of individuality to the collectivity, as in an authoritarian regime.

This democratic failure may be seen in the preference model as well. In theory, an impartial spectator properly computes utilities and suggests which policy will best improve social welfare. In practice, utilitarian policies suppose that some experts have made the utility computation in the name of the other individuals. In the ideal preference model, the computation also supposes they launder individual utilities. What is the actual legitimacy of such experts? How can we be sure that such an utilitarian elite may desire and succeed in promoting the good for the mass of people? The problem would not arise if a normative demarcation of the scope of expertise were possible. Under such a demarcation, experts could be contracted within a well-defined area in which they are responsible solely for making factual observations and computations. The citizens would have previously decided that the policy should be utilitarian, such that the experts would use utilitarian models to derive their prescriptions. As a consequence, expert decisions based on utilitarianism would legitimately proceed from the will of individuals, although in fact implemented by an expert. The possibility of such a demarcation is, however, doubtful (Baujard 2013). In the end, this objection again recalls the issue of the lack of democracy in the elaboration of actual policy recommendations and the implementation of utilitarian policies. An alternative to utilitarianism hence supposes the introduction of theories of democracy within welfare economics, *e.g.*, following Sen's more recent ideas (Sen 2009).

Conclusion

In seeking to deepen her understanding of this subject, the reader will quickly find herself within a wide arena of impressively numerous primary and secondary references on utilitarianism and its criticism. Glover's introduction to and collection of utilitarian texts (Glover 1990), as well as his thematic bibliography (Glover 1990: 251–255), provide a useful guide within this otherwise daunting literature.

Unquestionably, utilitarianism has evolved and developed since its appearance in the 18th century. In order to retain the essential kernel of the theory—that what is important for society is that individuals have high utility—authors have sought creative responses to the

tensions raised by the opposition between positive and normative definitions of utility, and individual and collective definitions of welfare. Utilitarianism has consequently been permanently modified and refined. This constant adaptation has made utilitarianism into the most important ethical theory in the Anglo-Saxon world, and in particular within the evolution of welfare economics.

Whenever the theory has been modified to repel strong criticism, it has increased its theoretical stability while aggravating its practical fragility. It hardly appears possible to answer both the practical and ethical objections raised against the utilitarian—or at least welfarist—bases of welfare economics. The ethical criticism returns when the theory is to be applied and implemented in concrete settings, *i.e.*, when the formulation of actual policy recommendations is at stake. Conceived and reconstructed on the bases of the most refined versions of contemporary utilitarianism, a welfarist welfare economics may be bound to remain a beautiful but vain theory, as is indeed presaged by the news of the death of welfare economics. Considering the practical stakes, going beyond utilitarianism, and even beyond welfarism, in welfare economics, may now prove to be an absolute necessity.

References

- [1] ARNESON, R. J. Equality and equal opportunity for welfare. *Philosophical Studies* 56 (1989), 77–93.
- [2] ARROW, K. J. Some ordinalist-utilitarian notes on Rawls’s theory of justice. *The Journal of Philosophy* 70, 9 (1973), 254.
- [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] BAUJARD, A. Welfare economics. In *Handbook of the History of Economic Analysis*, G. Faccarello and H. D. Kurz, Eds., vol. 3. Edward Elgar Publishing Limited, Cheltenham (U.K.), 2012. forthcoming.
- [5] BAUJARD, A. Value judgments and economics expertise. Working Paper GATE L-SE, WP 1314, 2013. [url: <ftp://ftp.gate.cnrs.fr/RePEc/2013/1314.pdf>].

- [6] BENTHAM, J. *Introduction to the principles of morals and legislation*. Blackwell's political texts. Blackwell, Oxford, 1789. Intr. de W. Harrison, 1967.
- [7] BENTHAM, J. *Panoptique : Mémoire sur un nouveau principe pour construire des maisons d'inspection, et nommément des maisons de force*. Mille et une nuits, Paris, 1791. Trad. de C. Laval, 2002.
- [8] BLACKORBY, C., BOSSERT, W., AND DONALDSON, D. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. No. 39 in Econometric Society Monographs. Cambridge university Press, 2005.
- [9] CAILLÉ, A. De l'anti-utilitarisme. Anniversaire, bilan et controverses. *Revue du M.A.U.S.S.*, 27 (1er. sem. 2006).
- [10] COHEN, G. A. On the currency of egalitarian justice. *Ethics* 99 (1989), 906–944.
- [11] DRIVER, J. The history of utilitarianism. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., summer 2009 ed. 2009.
- [12] DWORKIN, R. *Taking rights seriously*. Harvard University Press, Cambridge: MA, 1978.
- [13] DWORKIN, R. What is equality? Part 1: Equality of welfare. *Philosophy and Public Affairs* 10 (1981), 185–246.
- [14] DWORKIN, R. What is equality? Part 2: Equality of resources. *Philosophy and Public Affairs* 10 (1981), 283–345.
- [15] FACCARELLO, G., AND KURZ, H. D. *Handbook of the History of Economic Analysis*. In 3 volumes. Edward Elgar Publishing Limited, Cheltenham (U.K.), Forthcoming.
- [16] FOUCAULT, M. *Surveiller et punir*. Tel. Gallimard, Paris, 1975.
- [17] GLOVER, J. *Utilitarianism and its critics*, 1990.
- [18] 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.
- [19] GRIFFIN, J. *Well-being: Its meaning, Measurement, and Moral Importance*. Clarendon Press, Oxford, 1986.
- [20] HARE, R. M. *Moral thinking: Its levels, method and point*. Clarendon Press, Oxford, 1981.

- [21] HARROD, R. F. Utilitarianism revisited. *Mind* 45 (1936), 137–156.
- [22] HARSANYI, J. C. Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy* 61, 5 (1953), 434–435.
- [23] HARSANYI, J. C. Cardinal welfare, individual ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63, 4 (1955), 309–321.
- [24] HARSANYI, J. C. Can the maximin principle serve as a basis for morality? A critique of John Rawls theory. *American Political Science Review* 59, 2 (1975), 594–606.
- [25] HASLETT, D. What is utility ? *Economics and Philosophy* 6 (1990), 65–94.
- [26] HAUSMAN, D. M. *Preferences, value, choice, and welfare*. Cambridge University Press, 2012.
- [27] JEVONS, S. *The theory of political economy*. MacMillan, Londres, 1871.
- [28] MILL, J. S. *L'utilitarisme*, Quadrige ed. Presses Universitaires de France, 1861. Trad. fr. par C. Audard, 1998.
- [29] MOORE, G. E. *Principia Ethica*. Cambridge University Press, Cambridge, 1903.
- [30] NOZICK, R. *Anarchy, state and utopia*. Basic Book, New York, 1974. trad. fr. par E. D'Auzac de Lamartine, *Anarchie, État, et utopie*, PUF, Paris, 1988.
- [31] NUSSBAUM, M. C., AND SEN, A. K. *The Quality of Life*. Clarendon Press, Oxford, 1993. Edition 2001.
- [32] PARFIT, D. *Reasons and persons*. Oxford University Press, Oxford, 1984.
- [33] 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.
- [34] RAWLS, J. Reply to Alexander and Musgrave. *The Quarterly Journal of Economics* 88 (1974), 633–655.
- [35] ROEMER, J. E. *Equality of opportunity*. Harvard University Press, Cambridge, 1999.

- [36] ROSEN, F. *Classical Utilitarianism from Hume to Mill*. Routledge, London, 2003.
- [37] SEN, A. K. The impossibility of a Paretian liberal. *Journal of Political Economy* 78, 1 (1970), 152–157.
- [38] SEN, A. K. Equality of what ? In *The Tanner Lectures on human Values*, S. McMurrin, Ed., vol. 1. Cambridge University Press, Cambridge, 1979. Rééd. in A. K. Sen, *Choice, Welfare and Measurement*, MIT Press, Cambridge, pp. 365–366, 1982.
- [39] SEN, A. K. Personal utilities and public judgements : Or what's wrong with welfare economics? *The Economic Journal* 89, 355 (septembre 1979), 537–558.
- [40] SEN, A. K. Utilitarianism and welfarism. *The Journal of Philosophy* 76, 9 (1979), 463–489.
- [41] SEN, A. K. *The idea of justice*. Harvard University Press, Cambridge (MA), 2009.
- [42] SIDGWICK, H. *The methods of ethics*. Macmillan, 1874.
- [43] SKYRMS, B. *Evolution of the social contract*. Cambridge University Press, Cambridge, 1996. ed. 1998.
- [44] SUMNER, L. W. *Welfare happiness and ethics*. Oxford University Press, 1996.
- [45] WALZER, M. *Spheres of justice. A defense of pluralism and equality*. Basic Books, New York, 1983. Trad. fr. par P. Engel : *Sphères de justice. Une défense du pluralisme et de l'égalité*, Seuil, Paris, 1997.
- [46] WILLIAMS, B. A critique of utilitarianism. In *Utilitarianism: For and Against*, J. Smart and B. Williams, Eds. Cambridge University Press, Cambridge, 1973, pp. 75–150.
- [47] YAARI, M., AND BAR-HILLEL, M. On dividing justly. *Social Choice and Welfare* 1 (1984), 1–24.