Utilitarian population ethics: a survey

Gregory Ponthiere[*] [**]

Abstract:

This essay is a survey of utilitarian criteria aimed at guiding what Parfit (1984) called Different Number Choices (i.e. choices affecting both people's number and identities). The emphasis is laid on two aspects of those criteria: their ethical foundations and their implications. Our analysis starts with total, average and critical-level utilitarianisms, against which numerous criticisms, such as Parfit's Repugnant Conclusion and Mere Addition Paradox, were formulated, so that alternative social welfare criteria, as the ones developed by Hurka (1983) and Ng (1986), might seem appealing. However, those criteria are not fully satisfactory, and, as most criteria considered here, they do not stand up to Naverson's (1967) critique, according to which social welfare cannot be increased or reduced if no existing person is affected. The difficulties resulting from taking Naverson's critique into account – and thus from considering the Actual Problem rather than the Genesis Problem – are then discussed. It is concluded that utilitarian population ethics might reach an impasse, which might be regarded either as resulting from contradicting intuitions, or as an illustration of utilitarianism's own limits. But those limits would be faced by any other consequences-based ethical theory in front of Different Number Choices.

[*] F.N.R.S. Research Fellow and C.R.E.P.P., University of Liege (Belgium).
[**] The author thanks Professor Mirrlees for his comments and suggestions.

1: Introduction

As Parfit (1984) underlined, a government's choices can be of three distinct kinds: Same People Choices (affecting neither the number of people, nor their identities), Same Number Choices (affecting the identities of people, but not their number), and Different Number Choices (affecting both people's number and identities)[1]. The third kind of choice, being the most common, is also the most difficult to make, because it involves the difficult choice of the optimal population size. The present survey is concerned with choices of that kind. A particularly interesting Different Number Choice is what Dasgupta (1988) called the Genesis Problem. The Genesis Problem consists of the choice of the optimal population size given that no people actually exist, or, in other words, given that all people are potential. Although the Genesis Problem is an unrealistic problem, it has been the subject of numerous contributions made by philosophers and economists for the last 30 years. This is more recently that population theorists have started to think about another, more realistic, problem (assuming some actual population), which Dasgupta called the Actual Problem.

In this essay, I shall mainly confine myself to presenting some of the main theoretical results of the optimum population literature. Indeed, the present survey, having no pretension to exhaustiveness, is mainly concerned with trying to find a "satisfactory" social well-being criterion aimed at guiding a government in its optimal population choice[2]. The difficulty of our inquiry naturally depends on how one defines a "satisfactory" social well-being criterion. A first requirement might be that such a criterion, aimed at dealing with the difficult ethical question of the optimal population size, exhibits "strong" ethical foundations. Indeed, social welfare criteria cannot be practical rules without strong ethical foundations: the lack of such foundations would make them arbitrary and irrelevant for public choices. Moreover, one might also require that a "satisfactory" criterion generates "reasonable" consequences, or, in other words, "ethically desirable" consequences, when applied to particular situations. That second requirement, as we shall see, might take different forms, all of them referring – at least to some extent – to what we could call our "ethical beliefs" or our "intuitions": what we expect from a social welfare criterion is that it does not lead us to ethically "counter-intuitive" implications in terms of population size. The extent to which it is more or less possible to satisfy our second requirement naturally depends on what we mean by "counter-intuitive" consequences, and the difficulties resulting from the severity of such an "intuitions-based" requirement will be discussed latter in this paper.

Throughout this survey, I will remain within utilitarian ethics, so that I will confine myself to presenting utilitarian social well-being criteria. In other words, I will focus on social welfare criteria or functions (SWF), which are a particular subset of social well-being criteria. Following Kolm's (1969) terminology, our inquiry consists of looking for a "satisfactory" social welfare function of the "pragmatical" kind (representing the ethical preferences of a public agent choosing states of affairs), and not of the "categorical" kind (describing the society's preferences). Therefore this survey aims at trying to assess the adequacy of various SWF for optimal population choices, in the light of two requirements: ethical foundations and "reasonable" consequences. This survey might be regarded as an attempt to try to use those two requirements in order to "discriminate" between competing SWF, or, following Mackie's (1977) words, in order to try to overcome what he considered to be one of the "indeterminacies" in utilitarianism[3].

---

[1] See Parfit (1984), p. 355-356.

[2] More precisely, sections 2 to 8 will be concerned with the Genesis Problem, while sections 9 to 10 will deal with the Actual Problem.

[3] See Mackie (1977), p. 125-129.

Our choice to remain within the utilitarian tradition might be justified by the fact that most contributions in the optimal population literature were made within utilitarianism[4]. However, I think that such a choice is not too restrictive, because any consequentialist ethical theory (or, more generally, any consequences-based ethical theory) that tries to solve the optimum population problem might face, at least to some extent, the same difficulties as the ones faced by a purely consequentialist theory as utilitarianism[5]. It might well be true that utilitarianism, as many of its critics underlined, imposes restrictive constraints on the informational basis used for public choices. However, our present discussion is not about utilitarianism's informational basis, but about its combinatory structure. Therefore any consequentialist (or consequences-based) ethical theory, which must also be based on some combinatory structure, may also gain from theoretical "progress" made in the utilitarian optimal population inquiry.

Using utilitarian criteria (and thus welfarist criteria) as well-being criteria implies that we make the implicit assumption of welfarism[6]. It should be stressed, as Sen (1984) did, that the concept of "utility" might take three different senses: pleasure-happiness, desire-fulfilment, and the numerical representation of preferences. Most of the arguments developed here use the first or the second meaning of "utility"[7]. In order to be able to maximize SWF, we assume that individuals' utilities are perfectly measurable and that interpersonal comparisons of utility are possible. Moreover, we also assume, for convenience, that individuals are identical, and enjoy a positive but declining marginal utility of consumption. Furthermore, we will only consider the maximization of a SWF under a complete certainty[8].

The present survey on utilitarian population ethics is organized as follows. Firstly, it focuses on arguments for and against the two best-known utilitarian criteria: total utilitarianism (TU) (section 2) and average utilitarianism (AU) (section 3). Strong criticisms against those criteria lead us to look for an alternative SWF aimed at choosing the optimal population when no current population exists (Genesis Problem). An interesting alternative seems to be critical-level utilitarianism (CLU), developed by Blackorby and Donaldson (1984). However, as discussed in section 4, it exhibits the same weaknesses as total and average utilitarianisms, so that an "incomplete" version of CLU (critical-band utilitarianism, CBU), discussed in section 5, has been developed by Blackorby, Bossert and Donaldson (1996). Section 6, focussing on Parfit's (1984) *Reasons and Persons*, constitutes an attempt to summarize the various criticisms against TU, AU and CLU/CBU, and is aimed at clarifying the requirements made on a "satisfactory" SWF, called "Theory X" by Parfit. However, as we shall see, Ng (1989a) questioned the logical possibility of Parfit's "Theory X". Candidate SWFs are discussed in section 7 (Hurka, 1983) and section 8 (Ng, 1986). But those criteria are

---

[4] According to Sen (1984), utilitarianism is the combination of three things: firstly, welfarism, requiring that the only information used to assess the goodness of a state of affairs is the utility information associated with that state; secondly, consequentialism, requiring every choice to be determined by the goodness of the consequent state of affairs; thirdly, sum-ranking, requiring the goodness of a state to be assessed by looking at the sum of all the utilities in that state.

[5] Sen (1987) underlined the difference between a consequentialist theory (based on consequences only) and a consequences-based theory (based, among other things, on consequences). Sen argued in favour of an ethical theory based on consequences, but not strictly consequentialist.

[6] See footnote 4.

[7] However, the third meaning of "utility" will intervene at some point in the debate (regarding Sen's interpretation of Harsanyi's derivation of average utilitarianism).

[8] Given that the present survey focuses exclusively on the normative side of the maximization problem, no attention has been paid to the explicit formulation of the constraints to which the maximization problem is subject (except in the case of the Relative Repugnant Conclusion, see *infra*). When considering the "reasonableness" of the implications of various SWF, only a sketch of why such implications occur will be provided. But given that the implications resulting from the use of a particular SWF rather than another are most often straightforward, the omission of the constraints does not raise particular difficulties.

not fully satisfactory. Naverson's (1967) famous critique is then presented in section 9. Naverson argued against traditional utilitarian criteria that they were wrong when they claim that social welfare could be increased by the addition of a person. According to Naverson, social welfare cannot be increased or reduced if no existing person is affected by the choice. Once Naverson's critique is accepted, the problem we have to deal with is no longer the Genesis Problem, but the Actual Problem, which assumes the existence of actual people. Section 9 also analyses how traditional social welfare criteria deal with the Actual Problem. As we shall see, most of them do not stand up to Naverson's critique. Hence section 10 presents briefly some solutions to Naverson's critique. The most intuitive one is based on what Broome (1994) called the Constituency Principle. However, that principle leads to contradictions. One way to avoid them – and to satisfy Naverson's intuitions – might be based on Dasgupta's (1988) population theory, which provides a solution to the Actual Problem. However, that solution to Naverson's critique exhibits some difficulties to satisfy our two requirements: "strong" ethical foundations and "reasonable" consequences. Hence, as it is argued in the conclusion (section 11), the optimal population problem, whatever we consider that there are actual people (Actual Problem) or not (Genesis Problem), seems to remain unsolved so far. One explanation for our unsatisfactory results might be that our "intuitions" or ethical beliefs might be too demanding – or too self-contradictory – in order to serve as a guide to select the right SWF to make Different Number Choices in general, and to solve, in particular, the Genesis Problem or the Actual Problem. Another explanation might be that utilitarianism, as any ethical theory, cannot answer all thinkable questions.

2: Total utilitarianism

Regarding the optimal population choice, total utilitarianism suggests that the population size should be such that the sum of everybody's lifetime welfare is maximized. In a static framework, a total utilitarian SWF can be written:

$$W = Nu(c) \tag{1}$$

where N is the size of the population, while u(c) is the utility derived by an individual out of consumption c. Given our assumption of identical individuals, u(c) is the same for everyone in the society, and is also equal to the average utility prevailing in that society. In an intertemporal framework, a total utilitarian SWF can be written:

$$W = \sum_{t=0}^{\infty} N_t u(c_t) \tag{2}$$

where N_t is the size of generation t[9]. For simplicity, it is assumed that individuals live only one period of time, so that each period of time is associated with another generation of identical individuals.

It is often argued that total utilitarianism is the closest to what utilitarian pioneers considered to be utilitarianism. However, it should be stressed that neither Bentham nor Mill did explicitly express their preferences regarding the TU/AU debate. Indeed, the first thinker who made explicitly the distinction between total and average utilitarianism was Sidgwick

[9] Throughout this paper, I follow Sidgwick's (1874) interpretation of utilitarianism as "universalistic hedonism", so that I assume a zero "pure" discount rate. For a survey on the "pure" discount rate debate, see Ponthiere (2003).

(1874). But, as Birnbacher (1988) stressed, Bentham is often considered to support total utilitarianism implicitly, simply because of the addition of the factor "extent" in his famous *Introduction to the Principles of Morals and Legislation* (1789). Indeed, Bentham wrote[10]:

> "To a number of persons, with reference to each of whom the value of a pleasure or a pain is considered, it will be greater or less, according to seven circumstances […]: (1) its *intensity*; (2) its *duration*; (3) its *certainty* or *uncertainty*; (4) its *propinquity* or *remoteness*; (5) Its *fecundity*; (6) its *purity*; (7) its *extent*; that is, the number of persons to whom it *extends*; or […] who are affected by it."

Should we conclude from that quotation that Bentham supported total utilitarianism? According to Birnbacher (1988), we should not. Indeed, Bentham seems to take the number of people as given. In such a case, total and average utilitarianisms are equivalent. Therefore, Bentham, neglecting Different Number Choices, does not seem to choose between TU and AU. Regarding Mill, it is often argued, as in Gottlieb (1945), that Mill supported – at least implicitly – average utilitarianism (see *infra*). However, Sidgwick (1874), unlike Mill, introduced explicitly total and average utilitarianism, and argued in favour of the former as the best representation of utilitarianism as "universalistic hedonism". Indeed, Sidgwick wrote[11]:

> "Assuming, then, that the average happiness of human beings is a positive quantity, it seems clear that, supposing the average happiness enjoyed remains undiminished, Utilitarianism directs us to make the number enjoying it as great as possible. But if we foresee as possible that an increase in numbers will be accompanied by a decrease in average happiness or *vice versâ*, a point arises which has not only never been formally noticed, but which seems to have been substantially overlooked by many Utilitarians. For if we take Utilitarianism to prescribe, as the ultimate end of action, happiness on the whole, and not any individual's happiness, unless considered as an element of the whole, it would follow that, if the additional population enjoy on the whole positive happiness, we ought to weigh the amount of happiness gained by extra number against the amount lost by the remainder. So that, strictly conceived, the point up to which, on Utilitarian Principles, population ought to be encouraged to increase, is not that at which average happiness is the greatest possible – as appears to be often assumed by political economists of the school of Malthus – but that at which the product formed by multiplying the number of persons living into the amount of average happiness reaches its maximum."

Sidgwick clearly supported total utilitarianism. Numerous other utilitarian thinkers also prefer total utilitarianism to average utilitarianism, for various reasons[12]. According to Temkin (1992), one cannot be a utilitarian thinker if one is not a total utilitarian. Temkin argued that the superiority of TU is not a matter of plausibility, but rather a matter of "intrinsic appeal", of "intuition". A utilitarian, believing that utility is all that matters, must always prefer more of the Good to less of the Good. Therefore he must be a total utilitarian. This argument of "intuition" might seem artificial, but it reminds us that the utilitarian doctrine was initially aimed at competing (at the individual and the public levels) against religious beliefs and social conventions. One cannot outweigh religions and conventions without "appeal" or "intuition".

Therefore total utilitarianism seems to exhibit strong ethical foundations. Those foundations, as Dasgupta (1993) underlined, could also take an axiomatic form, based, for instance, on Blackorby and Donaldson (1984)'s work.

---

[10] Bentham (1789), republished in Mill, *Utilitarianism* (edited by Warnock), (1962), p. 65 (with original italics).
[11] Sidgwick (1874), p. 415-416.
[12] For instance, Birnbacher (1988) and Ng (1989a) are defenders of utilitarianism in its "total" form (see *infra*).

However, despite its "intrinsic appeal", total utilitarianism has been severely criticized[13]. The strongest criticism against total utilitarianism is probably Parfit (1984)'s Repugnant Conclusion[14]:

> "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living."

Parfit regarded that conclusion as "intrinsically repugnant". I personally find that conclusion repugnant. Total utilitarianism considers any state with positive average utility less desirable than a state in which average utility is smaller – but still positive – and the population is sufficiently larger. Hence total utilitarianism allows population size to substitute for any positive average utility, whatever the extent to which that level is low. Total utilitarianism would lead us to enormous populations, and, in an intergenerational framework, to a succession of enormous populations: as long as the net gain in total utility from a supplementary person is positive, it is always desirable to increase the population. Huge populations are not themselves repugnant: what is repugnant is the complete "quality-quantity" substitution that total utilitarianism allows.

However, Dasgupta (1988) argued, rightly in my view, that, although Parfit's conclusion is repugnant, its phrasing exaggerates its repugnancy: once we accept the Genesis Problem, there is nothing ethically repugnant about "quantities" of lives compensating for "qualities" of lives, as long as lives are worth being lived. Hence the repugnancy of the conclusion might also come from the problem asked.

Ng (1989a) went further, by arguing that the Repugnant Conclusion is compelling. According to Ng, if most people find Parfit's Repugnant Conclusion repugnant, this could be due to their "inability to understand the implication of large numbers" or to their "misplaced partiality"[15]. Although the first reason might seem acceptable, I am personally suspicious regarding the "misplaced partiality" motive. I think that the repugnancy, coming from the complete "quality-quantity" substitution, cannot be reduced to a result of humans' "misplaced partiality": if most human beings (except Ng) consider that Parfit's conclusion is repugnant, this, I believe, cannot be a simple expression of human's self interests.

Anticipating "plausibility" criticisms, Parfit (1984) argued that his scenario is not "deeply" impossible. However, Birnbacher (1988) argued that, in an intergenerational framework, the "quality-quantity" substitution holds only in principle. Birnbacher argued that resources pressures and political tensions make that substitution extremely limited in reality, beyond a couple of generations: one cannot imagine an infinite succession of large poor generations. However, I personally consider Parfit's scenario to be imaginable, even in an intergenerational framework. Indeed, Parfit's argument does not depend on any extreme assumption (no citizen having "extreme" preferences). Therefore, despite the numerous counter-arguments, I think that Parfit's Repugnant Conclusion – whatever it is "intrinsically repugnant" or only "repugnant" – remains a strong argument against total utilitarianism[16].

---

[13] This section will focus on criticisms against total utilitarianism in a static context, but those criticisms also hold in an intergenerational context.

[14] Parfit (1984), p. 388. More precisely, following Michel and Pestieau (1998), one could define Parfit's Repugnant Conclusion as the "Absolute Repugnant Conclusion", that is, "[a solution] implying a very large population of individuals consuming each a minimal amount of resources" (Michel and Pestieau (1998), p. 27). That Absolute Repugnant Conclusion should be distinguished from what Michel and Pestieau called the "Relative Repugnant Conclusion" (see *infra*).

[15] Ng (1989a), p. 242.

[16] Regarding the possibility of "avoiding" Parfit's Repugnant Conclusion by considering parenting rights, see Hammond (1988).

As Michel and Pestieau (1998) underlined, one could also argue that total utilitarianism does not only lead to what they called Parfit's "Absolute Repugnant Conclusion", but also to what they called the "Relative Repugnant Conclusion", or the "Marginal Repugnant Conclusion". Michel and Pestieau defined the Relative Repugnant Conclusion as follows[17]:

> "The *relative* repugnant solution occurs when any increase in the resources brought by each additional individual leads to a drop in equilibrium consumption."

Therefore, according to Michel and Pestieau (1998), total utilitarianism not only leads to an infinite population of individuals with very low consumption (the Absolute Repugnant Conclusion), but also to a solution where any increase in an individual's income generates a drop in the optimal level of consumption. In my view, that second Repugnant Conclusion, although this has been less debated than the first one, is as repugnant as the first one: what the Relative Repugnant Conclusion suggests is a picture of a growing economy with a decreasing individual consumption. This does not correspond to an intuitively desirable solution, so that the Relative Repugnant Conclusion definitely casts some doubts on the adequacy of total utilitarianism as an appropriate social welfare function.

Other criticisms were made against total utilitarianism. For instance, Cowen's (1989) "Methuselah's Paradox" used an analogy between the optimal population for a society and the optimal structure of an individual's life in order to attack sum-ranking principles (and thus total utilitarianism). Cowen developed his "Methuselah's Paradox" in order to show that the Repugnant Conclusion is generated by the sum-ranking component of total utilitarianism. As Cowen underlined, the main difference between Parfit's Repugnant Conclusion and Methuselah's Paradox is whether utilities are summed across additional individuals or across additional life-years of an existing individual. Methuselah's Paradox was expressed as follows[18]:

> "For any possible ecstatically happy and profound life of, say, 200 years, we can imagine another, much longer life which will welfare-dominate it simply by multiplying many years of epsilon utility".

According to Cowen, that result is obtained if we assume that additional years of life have some positive value. However, as Cowen argued, most people would prefer 200 years of a very happy life to a much longer life made of very small yearly utilities. Therefore we reach a paradox. Hence Cowen claimed that[19]:

> "If the value of a life cannot be represented as a sum of constituent units, then perhaps the value of a society cannot be so represented either."

However, that analogy exhibits some limitations. Indeed, as Cowen himself acknowledged, it is likely that the interconnections between the utility levels in different periods of an individual's life are stronger than the relations between the utilities of different persons in a society, so that assuming additive separability of utilities across time might be less attractive than the assumption of additive separability of utilities across individuals[20].

---

[17] Michel and Pestieau (1998), p. 27. Michel and Pestieau considered a model where the resources constraint is defined as follows: each newcomer generates some "costs" (consumption and education), but also brings some fixed amount of income, so that the resources constraint in their model is that the total number of people times their associated "costs" should not be higher than people's "contributions" plus a fixed stock of resources.

[18] See Cowen (1989), p. 37. Cowen assumed a zero rate of time preference.

[19] Cowen (1989), p. 38.

[20] Such a point of view was also expressed by Broome (1992).

Moreover, Ng (1989b) severely criticized Methuselah's Paradox. According to Ng, rationality requires that we choose the longest life, "provided that the uncertainty-discounted and impatience-adjusted sum of utilities is larger"[21]. Ng offered three explanations regarding people's "myopia". Firstly, people might misunderstand the dilemma: by his choice of a shorter life, the individual would not be short of time to fulfil his obligations towards future people, because it is assumed that the person's choice does not influence anyone else's happiness. Secondly, Ng suggested that human beings might be "genetically and/or culturally programmed" to desire the longest possible (thinkable) life, but not a life of a thousand of years. Thirdly, Ng argued that people might be unable to appreciate the significance of large numbers correctly. All those reasons might explain, according to Ng, why most people would exhibit inconsistent preferences in favour of a 200 years life. Although I am not entirely convinced by Ng's defence of total utilitarianism, I think that Cowen's critique might lack some weight, because of the inadequacy of the time-separability assumption. Therefore, Parfit's Repugnant Conclusion seems to be the strongest criticism against total utilitarianism.

To summarize, total utilitarianism seems to exhibit strong ethical foundations, in the sense that it seems to be the closest to the utilitarian goal: the promotion of the greatest happiness of the greatest number. However, total utilitarianism, leading to Parfit's (Absolute) Repugnant Conclusion, does not seem to satisfy our second requirement of "reasonable" consequences. Moreover, as Michel and Pestieau (1998) argued, total utilitarianism would also lead to the Relative Repugnant Conclusion, which does not seem to be intuitively appealing.

3: Average utilitarianism

Under average utilitarianism, what the public agent maximizes is the average welfare of all the people. In a static framework, average utilitarianism implies the maximization of the following SWF:

$$W = u(c) \tag{3}$$

where u(c) corresponds to the utility derived by any (identical) individual in the society, and thus corresponds to the society's average utility. In an intertemporal framework, average utilitarianism can be defined in two ways: on the one hand, as the maximization of the intertemporal sum of each generation's average utility; on the other hand, as the maximization of the average utility of all the people who will ever live, that is, the ratio of the intertemporal sum of each generation's total utility to the total number of people who will ever exist[22]. These two definitions correspond to the maximization of equation (4) and (4') respectively:

$$W = \sum_{t=0}^{\infty} u(c_t) \tag{4}$$

$$W = \sum_{t=0}^{\infty} N_t u(c_t) \Big/ \sum_{t=0}^{\infty} N_t \tag{4'}$$

If we remain within a static framework, for a given population, total utilitarianism (equation (1)) and average utilitarianism (equation (3)) lead to the same optimal policies. However,

[21] Ng (1989b), p. 47.
[22] On those two definitions, see Dasgupta (1993), chapter 13.

once the population size is chosen, those criteria lead to very different optima: while, under TU, the population size will be increased until no change in total welfare results, AU will lead to an optimal population size such that any increase in population would leave the average utility unchanged, according to equation (3). Under AU, a larger population is never used to compensate for a decrease in average welfare, as it is done under TU as long as total welfare is increased thanks to the addition of another person. Therefore average utilitarianism will generate a smaller optimal population, and thus does not lead to Parfit's Repugnant Conclusion (neither Absolute nor Relative)[23].

According to Gottlieb (1945), average utilitarianism was supported by Mill. Birnbacher (1988) also underlined that Mill seemed to support average utilitarianism implicitly, because Mill argued in favour of the control of births on the grounds of social welfare, which might be interpreted as a sign of some reliance on AU rather than TU. Indeed, Mill (1859) wrote[24]:

> "The fact itself, of causing the existence of a human being, is one of the most responsible actions in the range of human life. To undertake this responsibility – to bestow a life which may be either a curse or a blessing – unless the being on whom it is to be bestowed will have at least the ordinary chances of a desirable existence, is a crime against that being. And in a country either over-peopled, or threatened with being so, to produce children, beyond a very small number, with the effect of reducing the reward of labour by their competition, is a serious offence against all who live by the remuneration of their labour. The laws which, in many countries on the Continent, forbid marriage unless the parties can show that they have the means of supporting a family, do not exceed the legitimate powers of the State: and whether such laws be expedient or not (a question mainly dependent on local circumstances and feelings), they are not objectionable as violations of liberty."

Mill's support of the control of births might be interpreted as a preference for average utilitarianism over total utilitarianism. However, it should be stressed that such an interpretation is nothing more than an interpretation. Mill never referred explicitly to the maximization of average utility: as I stressed above, this is Sidgwick (1874) who was the first who introduced explicitly the distinction between total and average utilitarianism (see *supra*).

The main criticism against average utilitarianism, whatever we assume a static or a dynamic framework, is, as Temkin (1992) argued, its lack of ethical foundations. At first glance, average utilitarianism looks like an *ad hoc* solution to avoid total utilitarianism's weaknesses. However, it is often forgotten that ethical foundations to AU were provided by Harsanyi's (1955) "veil of ignorance" argument, based on the concept of "ethical preferences"[25]. In short, Harsanyi's (1955) axiomatic argument, in a static framework, goes as follows[26]. Suppose a rational individual[27], who must choose the optimal infinite consumption

---

[23] Under the assumption that, for some population size, the output available for consumption starts increasing less than proportionally to the population. I do not analyse here the properties of average utilitarianism in a dynamic context. Whatever the resources constraints used, AU defined as equation (4') will never lead to the Repugnant Conclusion (neither Absolute, nor Relative), but rather to an extremely low level of intertemporal population (see *infra*). AU defined as equation (4) would lead to a higher intertemporal population than in (4'), but still much lower than under total utilitarianism, because equation (4) suggests that each generation should be of an extremely small size. AU defined as equation (4) would generate neither Absolute nor Relative Repugnant Conclusion (the last result can be obtained by assuming α equal to 1 within Michel and Pestieau (1998)'s dynamic model, p. 30-31).

[24] Mill (1859), p. 242.

[25] Harsanyi firstly formulated his argument in another article (1953).

[26] Regarding the extension of Harsanyi's argument to a dynamic framework, see Dasgupta and Heal (1979). The difficulties generated by that extension in an infinite intertemporal context are discussed below. However, it should be stressed that an extension of Harsanyi's derivation to a dynamic context can only provide an ethical justification for the definition of average utilitarianism as the maximization of the average utility of any ever existing individual (equation (4')). However, as Dasgupta (1993) underlined, no obvious ethical foundation can be found for AU as defined in equation (4).

stream for the society. Suppose that he puts himself under a "veil of ignorance", that is, he forces himself to forget his social position, so that he faces a choice under uncertainty. Facing such an uncertainty, the individual will choose the infinite consumption stream maximizing his expected utility. According to Harsanyi, the individual's "ethical" preferences, in contrast with his "subjective" preferences, must be "impartial" and "impersonal", so that the individual must assign equal probabilities to all the potential social positions[28]. That probability being equal to the inverse of the population size, what the individual maximizes (his expected utility) corresponds to the average utilitarian criterion[29].

Numerous criticisms were formulated against Harsanyi's (1955) argument[30]. First of all, one might wonder, as Barry (1989), why moral judgements should necessarily be made behind a "veil of ignorance". In Barry's words[31]:

> "No adequate reason has ever been given (by Harsanyi or anybody else) for identifying moral judgements with those made by someone trying to maximize his own prospects from behind a veil of ignorance."

However, Broome (1991) considered that critique to hold only for Harsanyi's (1953) first argument, based on what Broome considered to be strong philosophical assumptions, but not for Harsanyi's axiomatic derivation (1955), which, according to Broome, requires only weak philosophical assumptions, and where the "veil of ignorance" does not seem to play the same role as in Harsanyi's original non-axiomatic argument[32].

Harsanyi's (1955) original axiomatic derivation of average utilitarianism was also criticized severely by Rawls (1971) in his *Theory of Justice*. Firstly, Rawls argued that there does not seem to be objective grounds in the initial situation for assuming that one has an equal chance of being anybody. One justification could be to suppose that the individual forms his probabilities by following Laplace's Principle of Insufficient Reason. However, in his own theory, Rawls considered that the ignorance behind the "veil" was such that the individual could not form any probability distribution. Therefore Rawls preferred the non-probabilistic maximin criterion. However, as Sen (1970) underlined, equi-probability in Harsanyi (1955)'s framework comes from the fact that Harsanyi required "ethical" preferences to be "impersonal", and defined his "impersonality" in terms of "as if equi-probability". Thus if we adopt Harsanyi's definition of ethical preferences, we must also adopt the resulting "equi-probability". Moreover, Rawls also argued that the expectations

---

[27] "Rational" in the sense that the individual's preferences satisfy the Von-Neumann-Morgenstern axioms of rational behaviour under risk, or the Marschak postulates equivalent to them (complete ordering among all prospects, continuity, sufficient number of non-indifferent prospects, equivalence of mixture of equivalent prospects).

[28] As Rawls (1971) mentioned, one could justify "equi-probability" by assuming that the choosing party follows Laplace's Principle of Insufficient Reason, according to which, when one has no evidence at all, one assigns equal probability to each possible case.

[29] Under the assumption of identical individuals, Harsanyi's derivation would lead us to equation (3) in a static framework. The extension of Harsanyi's original demonstration to a finite dynamic context does not raise particular difficulties. However, as Dasgupta and Heal (1979) underlined, extending that derivation in an *infinite* intergenerational context generates additional complications. Indeed, in such a framework, there are an infinite number of potential social positions, so that, as Dasgupta and Heal noticed, it is difficult for the choosing party to assign equal probabilities to all the possible positions. Therefore, as Dasgupta and Heal underlined, talking about "equi-probability", when the future is infinite, makes no sense: one cannot have a uniform probability distribution defined over the infinite set of integers. As Dasgupta (2001) argued, the only way to give sense to "equi-probability" in this context would be to assume that the probability of extinction over the indefinite future is unity.

[30] Throughout this section, I shall confine myself to criticisms against Harsanyi's original axiomatic argument in a static framework. But those criticisms also hold in an intergenerational context.

[31] Barry (1989), p. 78-79, quoted in Broome (1991), p. 56-57.

[32] On that issue, see Broome (1991), chapter 3.

resulting from Harsanyi's rationale were spurious for two reasons. Firstly, those expectations are not based on one system of aims, in the sense that the individual computes the expected utility of being everybody, by taking each individual's systems of ends. Hence, Rawls argued that such expectations are meaningless. A second, related argument is that it is impossible for an individual to assess the value of each of the "total circumstances" of others, without any reference to his own conception of the Good. Hence, according to Rawls, an individual behind a "veil of ignorance" cannot assess the value of each social position. I personally think that the two latter criticisms – unlike the first one – might severely weaken Harsanyi's axiomatic derivation of average utilitarianism.

Another objection made to Harsanyi's analysis was expressed by Diamond (1967) and Sen (1970), who argued against the adequacy of the expected utility axioms for social decision-making. According to Diamond and Sen, the independence axiom does not allow the impartial individual to take into account the distributive aspects of the resulting outcomes. However, in social choices, one would like to know more than the mathematical expectation of welfare under a veil of ignorance, but also the precise distribution of expected welfare over individuals. In other words, one could expect from the choosing party to take into account in its choices not only the expected utility under "veil of ignorance", but also the whole distribution of expected welfare across people. In order to emphasize that point, Diamond (1967) developed the example of a society composed of two individuals A and B, which faces a choice between two alternatives α and β, with two equally possible states of nature, θ_1 and θ_2. The utility levels under each contingency are given in the following table[33]:

	If θ_1 occurs	If θ_2 occurs
Alternative α	$u_A = 1$, $u_B = 0$	$u_A = 1$, $u_B = 0$
Alternative β	$u_A = 1$, $u_B = 0$	$u_A = 0$, $u_B = 1$

As Diamond argued, a SWF resulting from Harsanyi's axiomatic derivation would be indifferent between the two alternatives α and β. However, it seems obvious that, from a social point of view, alternative β is more desirable than alternative α – and even "strictly preferable", according to Diamond –, because under alternative β each individual has the same expected utility equal to ½, while, under alternative α, the expected utility of individual A is 1 and the expected utility of individual B is 0. Hence alternative β seems more socially desirable than alternative α because it is more equitable. However, the expected utility theory would not discriminate between those two alternatives, because it has no distributive considerations: the only relevant information it requires is the total expected utility under each alternative (that is equal to 1 under each alternative). In my view, Diamond and Sen were definitely right when they argued that social choices, as opposed to individual choices, require a broader informational basis than the mere total expected utility: the distribution of expected welfare among the society might also be crucial for making public choices.

As Weymark (1991) noticed, Sen (1986) also criticized Harsanyi's framework for not providing an axiomatization of utilitarianism, but rather a representation theorem of the underlying social preferences. Sen's criticism is based on the idea that Harsanyi's concept of utility is not independent of the concept of preferences. In other words, Sen suspects Harsanyi to use a concept of utility taking the sense of "numerical representation of preferences". However, as Sen and others noticed, that third meaning of utility cannot constitute a basis for utilitarianism. Indeed, as Weymark (1991) rightly emphasized, utilitarianism requires the social ordering to be derived from the social welfare function, and not the inverse. Moreover,

[33] See Diamond (1967), pp. 765-766.

under the third meaning of utility, utility is only ordinally measurable but not comparable: thus it cannot be used as a basis for utilitarianism, requiring cardinality and interpersonal comparisons of utilities. However, according to Weymark, there seem to be some "clues" in Harsanyi's work that his concept of utility represents more than preferences only. Therefore, according to Weymark, Sen's interpretation of Harsanyi's theorem as a representation theorem of the underlying social preferences might not be completely justified[34].

Another powerful criticism against average utilitarianism does not concern its controversial ethical foundations provided by Harsanyi, but rather the counter-intuitive consequences resulting from its use. Indeed, Parfit (1984), through his Mere Addition Paradox, emphasized an absurd implication of the average utility principle. Indeed, average utilitarianism would consider that a situation A+, resulting from the Mere Addition[35], to the current population, of a population of people whose lives are worth being lived but who enjoy an average utility lower than the initial population's average, is worse than the initial situation A (without Mere Addition). This is counter-intuitive, because the addition of supplementary lives that are worth being lived, and that do not affect the initial population's welfare, cannot make the society worse off. This is, however, what average utilitarianism considers. That counter-intuitive implication of AU also holds in intergenerational contexts. Indeed, AU makes the desirability of an additional life today depend on past people's welfare. If we suppose, for instance, that people very "distant" in the past enjoyed a very high welfare, then the birth of any child today (equivalent to the Mere Addition of a person) would not be prescribed by AU, because it would lower the average utility. But, as Parfit underlined, lowering the average utility by Mere Addition cannot be regarded as bad. This is, however, what average utilitarianism does.

Moreover, as Hammond (1988) showed, if we consider long dead people not to be members of current or future society, then average utilitarianism becomes dynamically inconsistent. Indeed, as Broome (1992) argued, if we exclude people very "distant" in the past from our calculations, because we consider that their welfare should not influence our current choices, then the value of a policy depends on when it is evaluated, because that time determines which persons are included in the calculations.

Furthermore, Broome (1992) underlined that average utilitarianism would consider the death of a person having a positive utility level, but lower than the average, to be socially desirable. In the light of his criticisms and of Parfit's Paradox, Broome argued that the problem with AU was that it is not strongly separable between possible people: that principle makes the value of a person's existence depend on the welfare of other people. According to Broome, such a property seems to be difficult to defend on ethical grounds.

In conclusion, the ethical foundations of average utilitarianism provided by Harsanyi seem questionable. Rawls's criticism of the meaning of the expression derived by Harsanyi, and the Diamond-Sen's criticism of the use of expected utility theory for social choices seem to weaken Harsanyi's theoretical result seriously. Moreover, as Parfit (1984) and Broome (1992) underlined, average utilitarianism might lead to counter-intuitive population policies. Attempts to avoid Parfit's Paradox, as Hammond (1988) showed, could make AU time-inconsistent. Therefore average utilitarianism does not seem to satisfy our two requirements for a "satisfactory" social welfare criterion for optimal population choice.

[34] For a complete treatment of the "Sen-Harsanyi debate", see Weymark (1991).

[35] According to Parfit (1984, p. 420), there is *Mere Addition* when; "in one of two outcomes [we compare], there exist extra people (1) who have lives worth living, (2) who affect no one else, (3) whose existence does not involve social injustice".

4: Critical-level utilitarianism

As an alternative to total and average utilitarianisms, Blackorby and Donaldson (1984) developed what they called the critical-level utilitarianism (CLU), where the optimal population is the one that maximizes the sum of the differences between individual utilities and a constant non-negative critical level of utility α[36]. According to critical-level utilitarianism, adding a person to a population is good only if that person's utility is higher than the critical level α. In a static framework, critical-level utilitarianism implies the maximization of:

$$W = N\left(u(c) - \alpha\right) \tag{5}$$

while in a dynamic framework, CLU leads to the maximization of:

$$W = \sum_{t=0}^{\infty} N_t\left(u(c_t) - \alpha\right) \tag{6}$$

where the critical level α is assumed to be constant over time.

The critical level α allows a "compromise" between total and average utilitarianisms[37]: we partially avoid AU's undesirable effects of considering additional lives profitable only when a utility higher than the average is associated with them, because the addition of a person remains socially profitable only if his utility is higher than α.

Moreover, the critical level allows us to avoid Parfit's Repugnant Conclusion: it is socially optimal to increase the population until the additional person's utility is equal to α, but not further. Furthermore, it should be stressed that critical-level utilitarianism might also allow us to avoid the Relative Repugnant Conclusion. Indeed, as Michel and Pestieau (1998) showed, a SWF with a particular non-constant critical level of utility, linked to the marginal contribution of each added person to the total resource constraint of society, avoids the Relative Repugnant Conclusion[38]. Indeed, as Michel and Pestieau showed, a particular critical level of utility dependent on each additional individual's income allows us to avoid a solution where an increase in the income of a newcomer generates a drop in optimal consumption[39].

Therefore critical-level utilitarianism appears to be a reasonable compromise between total and average utilitarianisms. Moreover, some axiomatic derivations of CLU were developed by Blackorby, Bossert and Donaldson (1995) in an intertemporal framework. Indeed, they showed that critical-level generalized utilitarianism is the only family of population principles satisfying the axioms of Strong Pareto (implying Pareto indifference), anonymity, continuity, and independence of the utilities of the dead[40].

Another – involuntary – support to critical-level utilitarianism was provided by Broome (1992). According to Broome, a "reasonable" SWF should satisfy the requirements of strong separability between people and "impartiality". Strong separability between people, implying

[36] Generalized CLU can be obtained by applying a concave transformation on utilities and α.

[37] As Ng (1986, p. 375) noticed, TU and AU are two particular cases of CLU. TU is obtained when α is equal to zero, while AU is obtained when α is equal to "the maximum possible average (over persons) utility in all relevant social states".

[38] But a SWF with a constant critical level of utility would not avoid the Relative Repugnant Conclusion.

[39] However, Michel and Pestieau's CLU with a non-exogenous critical level of utility might be subject to Ng's (1986) remark that a SWF – and thus the value of the critical level α – should remain independent of any constraint (see *infra*).

[40] That last axiom allows history to matter to some extent, and thus allows us to avoid the Repugnant Conclusion.

that the "value" of an additional person should be independent of other people's welfare, excludes average utilitarianism. "Impartiality" here is the condition that identity does not matter; if a life is lived at a particular time, it does not matter who lives it. From those conditions and some others, Broome derived critical-level utilitarianism. Such a derivation might be regarded as an argument in favour of CLU. However, Broome regards CLU as an *ad hoc* solution, and hence concluded that individually plausible conditions might lead to implausible results, because they are in conflict with each other. This leads us to other criticisms oriented against critical-level utilitarianism[41].

According to Parfit (1984), critical-level utilitarianism does not really avoid the Repugnant Conclusion: whatever the situation, as long as the added person's utility is strictly higher than the critical level of utility, there always exists another situation with a larger population of people having a utility level closer to the critical level, but which is considered to be preferable, according to CLU. If the critical level is low, then such a conclusion is as repugnant as Parfit's Repugnant Conclusion[42].

However, as Broome (1992) emphasized, if the critical level is too high, then CLU exhibits the same undesirable property as AU: it considers that the addition of a person, whose life is worth being lived, is not socially profitable, because the utility associated with that additional life would be below the critical level. Hence there seems to exist a dilemma regarding the value of the critical level.

Moreover, Ng (1986) illustrated that critical-level utilitarianism might generate counter-intuitive rankings of social states when the average utility is strictly smaller than α. In Ng's words[43]:

> "For any $\alpha > 0$, we can imagine some set of constraints on resources, etc, such that comparisons of social states with average utility $v < \alpha$ may be involved. However, for the range $v < \alpha$, the Blackorby-Donaldson criterion gives completely counter-intuitive – if not unethical – rankings. For example, if $\alpha = 10$ utils, social state x ($v = 8$, $n = 10$) is ranked lower than social state y ($v = 3$, $n = 2$), where n is the number of individuals. However, in x, there are more people, with lives more worth living, than in y ($8 > 3 > 0$; $10 > 2$). […] Thus, since there are more persons in x, with worth-living lives [= with strictly positive utility], and more worth-living than those in y, it is natural to expect x preferred to y, as is preferred by both average and total utility maximization."

Ng's example stressed some counter-intuitive implications of critical-level utilitarianism as a criterion for optimal population choice.

Furthermore, regarding the choice of the critical level of utility, Ng (1986) underlined that it is tempting to choose it according to the capacity constraints of the planet, or according to some economic constraints. However, according to Ng, a "true" SWF, as an objective function to be maximized, should remain independent of any constraint, and should rather hold under various sets of economic constraints. This should also be the case of any component of a SWF, and thus of the value of the critical level of utility[44]. It follows from those criticisms that Ng is clearly sceptical regarding the choice of the critical level of utility.

---

[41] Throughout this section, I shall confine myself to criticisms against CLU in a static context, but those criticisms also hold in an intergenerational framework.

[42] On CLU, see Parfit (1984), p. 412-413.

[43] Ng (1986), p. 376.

[44] In particular, Ng's (1986) remark on the status of a SWF could hold against Michel and Pestieau's CLU with a non-exogenous critical level of utility (depending on the income of the newcomer). Ng would probably criticize Michel and Pestieau's SWF, by arguing that a "true" SWF – and thus the value of the critical level – should remain independent of any constraint. However, one could wonder whether any SWF satisfying Ng's requirement could ever avoid the Relative Repugnant Conclusion pointed out by Michel and Pestieau. But one could guess that Ng, who argued that the Absolute Repugnant Conclusion was not repugnant, does probably not regard the Relative Repugnant Conclusion as repugnant.

In conclusion, critical-level utilitarianism seems, at first sight, to be a quite attractive compromise between total and average utilitarianism. Critical-level utilitarianism seems to generate "better" consequences than TU or AU. Indeed, CLU can be regarded as immunized – at least to some extent – against the Repugnant Conclusion and the Mere Addition Paradox. However, the main criticism against critical-level utilitarianism attacks its central element: the critical level of utility. Indeed, as Broome argued, the critical level seems to be an *ad hoc* solution that has no strong justification. Moreover, as Ng showed, CLU with a critical level higher than the average utility might lead us to counter-intuitive implications. Hence the choice of the critical level of utility, which should be independent of any external constraint regarding the availability of resources, remains a controversial issue. Critical-band utilitarianism, discussed in section 5, might be regarded as an attempt to solve the difficult problem of the choice of the critical level of utility.

5: Critical-band utilitarianism

Following a suggestion from Parfit (1982), Blackorby, Bossert and Donaldson (1996) tried to solve the problem of the choice of the critical level of utility by defining an interval of critical levels $[\alpha^L, \alpha^H]$. Then, under that alternative assumption, critical-level utilitarianism becomes critical-band utilitarianism (CBU), or, according to Blackorby, Bossert and Donaldson's terminology, incomplete critical-level utilitarianism (ICLU). Under critical-band utilitarianism (or ICLU), the addition of a person is considered to be socially desirable (undesirable) if the added person's utility is above (below) the highest (lowest) critical level. However, if the added person's utility is within the interval defined by the two critical levels, that is, within the interval $[\alpha^L, \alpha^H]$, then critical-band utilitarianism does not recommend anything.

Therefore that alternative social welfare criterion provides rankings of states of affairs that are not necessarily complete: what is obtained is a quasi-complete ordering of states (rather than a complete ordering). Therefore, CBU (or ICLU) expresses the idea that sometimes it might not be possible to rank two states of affairs: in other words, it expresses the idea that the betterness relation might not be complete. This might be seen as a weakness of CBU. However, Blackorby, Bossert and Donaldson argued that such an incompleteness might be a strength: it leaves some room for moral ambiguity while other social welfare criteria, providing complete rankings, sometimes lead to counter-intuitive implications, such as Parfit's Repugnant Conclusion. Therefore the incompleteness is here seen as a way to avoid "unreasonable" consequences, and thus as a means to satisfy our second requirement.

Indeed, as Blackorby, Bossert and Donaldson (1996) underlined, critical-band utilitarianism allows us to avoid the Repugnant Conclusion, provided there is at least one positive critical level in the interval: only the addition of a person with a utility level higher than the upper bound of the interval will be regarded as socially desirable. The strength of the interval is that even a high upper bound does not lead to the Mere Addition Paradox. Indeed, additional lives with utility levels in the interval will be considered neither as socially desirable, nor as socially undesirable (unlike what was recommended under CLU). Therefore CBU seems to avoid the counter-intuitive consequences generated by total utilitarianism and average utilitarianism.

However, CBU does not seem to be immunized against Ng's (1986) criticism against CLU. Indeed, this is only under the assumption that the lower bound of the interval α^L is lower than the average utility that CBU does not rank state x (v = 8, n = 10) below state y (v = 3, n = 2). Provided the upper bound α^H is strictly lower than 8, x will be ranked above y, whatever the level of the lower bound of the interval. In the case in which the upper bound α^H

is equal to, or greater than 8, the two states x and y cannot be ranked against each other, provided the lower bound is not greater or equal to 8. However, if the lower bound is equal to 10, then y will be ranked above x, so that the criticism developed by Ng against CLU still holds under CBU for α^L higher than the average utility. Therefore CBU is not immunized against Ng's criticism.

Moreover, as Blackorby, Bossert and Donaldson (1996) underlined, critical-band utilitarianism does not solve all the problems regarding the choice of critical levels, especially when one abandons our simplified world of identical individuals. For instance, when one considers population's decisions, should one assign the same critical levels to men and women? Given that women have a longer life expectancy than men, and provided there is an asexual reproduction and a complete control of the sex of the baby before birth, then choosing the same critical levels for men and women would imply, from a social point of view, that only women should be born (this would maximize social welfare according to CLU or CBU). Therefore, according to Blackorby, Bossert and Donaldson, it might be justified to choose different critical levels for subsets of human beings that have different life expectancies but enjoy the same utility level per period. All this shows that the choice of critical levels remains a difficult task[45]. But other social welfare criteria do not provide more satisfactory results in those cases.

Furthermore, two other criticisms can be formulated against critical-band utilitarianism. Firstly, despite Blackorby, Bossert and Donaldson's axiomatic characterizations, CBU seems to lack strong ethical foundations. Broome (1992) was already sceptical regarding the use of a single critical level of utility. This is not surprising that a whole interval of critical levels is even more difficult to justify from an ethical point of view. Secondly, as Broome (1996) argued, it is not clear what it means that a state of affairs cannot be ranked against another. Does it mean that those two states of affairs are equally good? According to Broome, one might suspect CBU of achieving its nice consequences by some "cheating": either we consider that two states that cannot be ranked are equally good, and then we can choose between them in a random way; or if they are not equally good, then this means that one must be better than the other, and thus one could expect from a social welfare criterion to allow us to select the best of the two states. Therefore providing no complete ranking might be an attractive solution, because of its flexibility, but this might be regarded as a weak solution, in the sense that it fails at ranking alternatives that are not necessarily equally good.

To summarize, critical-band utilitarianism, despite its attractiveness, does not satisfy our requirement of strong ethical foundations: the introduction of an interval of critical levels seems even more *ad hoc* than the use of a single critical level. Regarding our second requirement ("reasonable" consequences), although CBU avoids Parfit's Repugnant Conclusion and Parfit's Mere Addition Paradox, it is not immunized against Ng's criticism. Therefore critical-band utilitarianism does not seem to be a solution to our inquiry.

6: Parfit's "Theory X"

At this stage, it might be worth summarizing the results of our inquiry. Sections 2 to 5 illustrated the difficulty to find a SWF having strong ethical foundations (our first requirement) and generating "reasonable" implications (our second requirement). Regarding the first requirement, total utilitarianism seems to be the best approach. However, it leads to Parfit's Repugnant Conclusion[46]. Average utilitarianism, whose ethical foundations are

[45] Blackorby, Bossert and Donaldson (1996) also discussed the critical levels to be chosen regarding policies affecting different animal species.

[46] Both "Absolute" and "Relative", according to Michel and Pestieau (1998)'s terminology.

controversial, suffers from its tendency to regard additional lives as socially unprofitable, even if they are worth being lived. Critical-level and critical-band utilitarianisms have more "reasonable" implications, but lack ethical foundations and hence seem arbitrary. Moreover, as Ng (1986) showed, CLU may violate our intuition in some cases. As we showed, critical-band utilitarianism is subject to the same criticism. Therefore neither CLU nor CBU seems to be fully satisfactory, even if we focus on our second requirement only.

Therefore the question remains: what would be a "reasonable" SWF for social choices involving population size? A first part of an answer to that question might be obtained by leaving aside – temporarily – our first requirement. One might then wonder what would be a social welfare criterion generating "reasonable" consequences, and also whether such a social welfare criterion exists.

A clarification of those requirements was provided by Parfit (1984) in his monumental *Reasons and Persons*[47]. According to Parfit, the "best" SWF, called "Theory X", should satisfy four requirements: it should solve the Non-Identity Problem, avoid the Repugnant Conclusion and the Absurd Conclusion, and solve the Mere Addition Paradox. While the Repugnant Conclusion (section 2), and the Mere Addition Paradox (section 3) were discussed above, the so-called No Identity Problem simply refers to the idea that one must find a social welfare criterion that allows us to compare states of affairs (or outcomes) where different people exist, that is, a social welfare criterion that allows us to make Different Number Choices[48]. Indeed, as Parfit underlined, there is no person far in the future whose existence does not depend of today's choices (or policies). Regarding the Absurd Conclusion, it should be stressed that it results from Parfit's attempt to avoid the Repugnant Conclusion and to solve the Non-Identity Problem by introducing an asymmetry about the value of quantity of the Good. More specifically, the Absurd Conclusion is the conclusion that we reach when we consider that quality has always value, but that quantity has value but cannot be above an upper limit in any period (so that an upper limit is placed on the quantity's positive value of the Good, but not on the quantity's negative value). The Absurd Conclusion was the following[49]:

> "In one possible outcome, there would exist during some future century both some population on the Earth that is like the Earth's present actual population, and an enormous number of other people, living on Earth-like planets that had become part of the Solar System. Nearly all of the people on these other planets would have a quality of life far above that enjoyed by most of the Earth's actual population. In each ten billion of these other people, there would be one unfortunate person, with a disease that makes him suffer, and have a life that is not worth living.
> In a second possible outcome, there would be the same enormous number of extra future people, with the same high quality of life for all except the unfortunate one in each ten billion. But this enormous number of extra future people would not all live in one future century. Each ten billion of these people would live in each of very many future centuries.
> On our view, the first outcome would be very bad, much worse than if there were none of these extra future people. The second outcome would be very good. The first would be very bad and the second very good even though, in both outcomes, there would be the very same number of extra future people, with the very same high quality of life for all except the unfortunate one in each ten billion."

[47] This section is mainly based on the Part 4 and the concluding chapter of Parfit (1984). It will focus on the choice of an optimal population in a static framework. Moreover, given that Parfit was not concerned with what Michel and Pestieau (1998) called the Relative Repugnant Conclusion (but only with the Absolute Repugnant Conclusion), this section will not consider the additional requirement of "avoiding the Relative Repugnant Conclusion". However, I shall come back on that additional requirement in sections 7 and 8.
[48] See Parfit (1984), p. 355-361.
[49] Parfit (1984), p. 410-411.

That conclusion is absurd, because the two outcomes differ only regarding the timing of the existence of the extra future people. It follows from that conclusion that one should not impose an upper bound on quantity's positive value[50].

Therefore, according to Parfit, "Theory X" should not only solve the Non-Identity Problem, avoid the Mere Addition Paradox and the Repugnant Conclusion, but also avoid the Absurd Conclusion. Parfit claimed that such a "Theory X" might exist.

However, Ng (1989a) argued that, from a logical point of view, "Theory X" cannot exist, because the Non-Mere Addition Paradox condition, with the Non-Antiegalitarianism condition[51], implicitly present in Parfit's work, implies the Repugnant Conclusion. Ng showed that result by means of the following three alternatives[52]:

> "A: 1 billion individuals with an average utility of 1 billion utils.
> A+: The same 1 billion individuals with exactly the same utility levels plus 1 billion trillion individuals each with 1 util (i.e., barely worth living).
> E: The same individuals as in A+ with a somewhat higher total utility but equally shared by all (i.e., each with, say, 1.01 utils)."

Ng's rationale goes as follows. According to the Mere Addition Principle (Non-Mere Addition Paradox), A+ is better than, or at least not worse than A. Moreover, Non-Antiegalitarianism suggests that E is better than A+, so that E is better than or at least not worse than A. However, given that each individual in E enjoys a utility level of 1.01, and thus a life being barely worth living, Ng concluded that saying that E is better than or at least not worse than A must still be regarded as an illustration of Parfit's Repugnant Conclusion. Ng concluded from that gamble that a "Theory X" satisfying both the Mere Addition Principle and Non-Antiegalitarianism and avoiding the Repugnant Conclusion cannot exist[53]. Therefore, one of those requirements would have to be given up. Ng underlined that the Non-Antiegalitarianism condition is an extremely reasonable requirement, satisfied by total and average utilitarianisms. Ng also noticed that if Parfit does not agree with the Non-Antiegalitarianism condition, then it cannot be claimed that his "Theory X" does not exist. However, in that case, it remains possible, according to Ng, to find a "Theory X", but that theory would then violate the Non-Antiegalitarianism condition[54]. According to Ng, such a "Theory X" would not be very attractive. Therefore, it seems difficult, according to Ng, to find an acceptable "Theory X", because Parfit's requirements, taken together, become extremely demanding. Ng's theoretical result is crucial, because it underlines the difficulty to find a social welfare criterion that would satisfy several "reasonable" requirements. Ng's "impossibility result" shows us how complex it is to find a SWF satisfying our second requirement ("reasonable" consequences). It follows from this that even if we get rid of our first requirement (strong ethical foundations), a "satisfactory" social welfare criterion for public choices involving population sizes seems difficult to find. One can only imagine the complexity of finding a SWF satisfying our two requirements. However, despite those

---

[50] Moreover, Parfit argued that one should not impose a limit on quantity's negative value: supplementary suffering cannot be socially ignored (see Parfit, 1984, p. 414).

[51] Non-Antiegalitarianism prevails "if alternative B has the same set of individuals as in alternative A, with all individuals in B enjoying the same level of utility as each other, and with a higher total utility than A, then, other things being equal, alternative B must be regarded as better than alternative A" (Ng, 1989a, p. 238).

[52] Ng (1989a), p. 240.

[53] However, as Ng (1989a) noticed, Parfit (1984, pp. 430-432) argued that rejecting the Mere Addition Paradox does not force us to accept the Repugnant Conclusion, because the Mere Addition Principle simply implies that A+ is not worse than (instead of "better than") A, and, moreover, the relation "not worse than" is not transitive. Such a defence was criticized by Ng (1989a, Appendix, pp. 252-253).

[54] On Parfit's attitude of rejection of the Non-Antiegalitarianism condition, and on – according to Ng – Parfit's "implicit" adherence to some form of it, see Ng (1989a, footnote 4 p. 239).

difficulties, some potential solutions were proposed as candidates to "Theory X". Those solutions, proposed by Hurka (1983) and Ng (1986, 1989a), are discussed in the next two sections.

7: Hurka's "variable value view"

In general terms, Hurka (1983) suggested that the contribution of an individual animal to the world's value is not constant, but depends on the number of other existing animals of his species. If the number of existing animals in his species is already high, then the added animal's "contribution" to the world's value is low, and his non-existence does not generate a big opportunity cost in terms of value. However, when the number of other existing animals in his species is very low, then his non-existence would be very costly, from a social point of view. Hurka called such a position the "variable value view", and attributed that view to Aquinas, Leibniz and Kant[55]. According to Hurka, the same view could be applied to human beings. Hence the value of the contribution of one individual to the society's value will depend on the number of existing people, and will be much higher if the initial human population is not large.

According to Hurka, total and average utilitarianism do not satisfy the "variable value view", and this might explain why those two social welfare criteria are unattractive when one has to deal with population choices. Hurka considered that average utilitarianism does not assign enough value to an additional life when the population is small: average utilitarianism rejects the addition of a supplementary person if this decreases the social average utility. But Hurka regarded this as undesirable, especially when the population is low. Regarding total utilitarianism, Hurka argued that it had undesirable consequences when the population is high (Parfit's Repugnant Conclusion), because it gives the same value to additional lives even when the population is already large. Hurka's "variable value view" would allow us to avoid Parfit's Repugnant Conclusion. Hence Hurka's "variable value view" would generate more "reasonable" consequences than total or average utilitarianism.

Therefore Hurka proposed two alternative criteria incorporating the "variable value view". Each of those criteria is such that an increase in the population, for a given average well-being, will always generate a decreasing positive effect on total welfare when the population becomes larger. That positive effect is always decreasing as the size of the population increases. The first version of Hurka's criterion (called V1) is such that, for a given population, a doubling of average well-being will always double the value of the whole population[56]. A second version (called V2) differs from the first in the sense that the value of a fixed increase in the average well-being, for a given population, will be positive but decreasing as the average welfare becomes higher. In a static framework, Hurka's V1 criterion can be written as follows:

$$W = \Phi(N)u(c) \tag{7}$$

where Φ(N) is a concave increasing function of N, while V2 can be written:

$$W = \Phi(N)\Delta\big(u(c)\big) \tag{8}$$

[55] According to Hurka, those thinkers applied that "variable value view" primarily to animals.

[56] Given that criterion V1 has become popularly used in the literature under the name, introduced by Ng (1986, 1989a), of number-dampened utilitarianism (NDU), I shall analyse V1's properties in greater depth in the next section dedicated to NDU only.

where Δ(.) is a concave increasing function. In an intertemporal framework, V1 and V2 can be written respectively:

$$W=\sum_{t=0}^{\infty}\Phi(N_t)u(c_t) \tag{9}$$

and

$$W=\sum_{t=0}^{\infty}\Phi(N_t)\Delta\big(u(c_t)\big) \tag{10}$$

Hurka's "variable value view" criteria seem to exhibit appealing properties. Indeed, it should be stressed that Hurka's criteria seem to be immunized – at least to some extent – against Parfit's Mere Addition Paradox. As Hurka pointed out, "variable value view" criteria do not forbid as many "mere additions" as average utilitarianism does, especially when the initial population is low[57].

Moreover, Hurka's criteria allow us to avoid Parfit's Repugnant Conclusion, because they assign a decreasing value to additional lives, thanks to the concavity of the function Φ(.) (see *supra*). However, it should be stressed that Hurka's criteria do not allow us to avoid what Michel and Pestieau (1998) called the Relative Repugnant Conclusion. Indeed, under Hurka's criteria, an increase in the income of a newcomer will still lead to a decrease in the optimal consumption[58]. Therefore Hurka's criteria would not allow us to avoid the Relative Repugnant Conclusion.

To summarize, if we focus only on Parfit's requirements for his "Theory X" (and thus if we do not consider the Relative Repugnant Conclusion), then Hurka's criteria, based on the "variable value view", seem to generate more "reasonable" implications than total or average utilitarianism, because of the decreasing value they assign to additional lives when the population increases. Hence these might be regarded, not as "perfect" candidates, but rather as "satisfactory" candidates to Parfit's "Theory X". However, Hurka's criteria fail to satisfy our first requirement. Indeed, what could justify that an additional person has varying social "value" according to the size of the already existing population? Hurka's criteria contradict completely Sidgwick's view of utilitarianism as universalistic hedonism[59]. As Broome (1992) argued, such an unfair treatment – and the lack of perfect separability between people – cannot be justified. All this might illustrate a "tension" between our two requirements: in order to have more "reasonable" consequences, one must give up ethical foundations.

8: Ng's number-dampened utilitarianism

Ng (1989a) not only showed the impossibility to find a "Theory X" satisfying a Non-Antiegalitarianism condition (see section 6), but he also developed his own solution to the problem, called "Theory X'", or "number-dampened" utilitarianism (NDU). Number-dampened utilitarianism is based on total utilitarianism, except that the value of adding a

[57] Regarding the Mere Addition Paradox, Hurka argued that criterion V2 seems to be better than criterion V1, because V2 makes the value of changes in the average well-being tend to zero as the average well-being becomes higher, so that V2 forbids fewer "mere additions" at high averages than V1 does.

[58] Indeed, Hurka's V1 and V2 criteria correspond respectively to critical-level utilitarianism and generalized critical-level utilitarianism with concave transformations on N, and with a constant critical level of utility equal to zero. Hence, under Hurka's V1 and V2 criteria, an increase in a newcomer's income would lead to a negative effect on the optimal level of consumption.

[59] "In Bentham's formula, 'everybody to count for one, and nobody for more than one'", Sidgwick (1874), p. 417.

person diminishes as the population increases. In other words, number-dampened utilitarianism suggests maximizing the product of the average social utility times an increasing concave function of the population size, exactly as what Hurka's V1 criterion suggests (see equations (7) and (9))[60]. According to Ng, number-dampened utilitarianism differs from Hurka's "variable value view" (V2), because NDU applies an increasing concave function to the population size only, and uses the same concave function on the population size to multiply any level of average social utility, while Hurka (1983)'s "variable value view" (V2) applies an increasing concave function on the product of the population size times the average social utility, and uses different concave functions as the average social utility differs[61].

Ng (1986, 1989a) underlined that number-dampened utilitarianism exhibits interesting properties: since it assigns a decreasing value to additional lives when the population increases and everything else remains constant, it avoids the Repugnant Conclusion[62]. Moreover, NDU solves the Non Identity Problem, without implying the Absurd Conclusion. Furthermore, NDU satisfies the Non-Antiegalitarianism condition. Regarding the Mere Addition Paradox, Ng argued that number-dampened utilitarianism avoids it in all the cases "regarded by most as more compelling", provided the chosen concave transformation function is not of extreme concavity[63]. Ng (1989a) also underlined that number-dampened utilitarianism is immunized against the criticism he developed against critical-level utilitarianism (and that also applies to critical-band utilitarianism): NDU never leads to counter-intuitive rankings.

However, several criticisms could be formulated against number-dampened utilitarianism. Firstly, although NDU seems relatively good at satisfying Parfit's requirements for his "Theory X", it should be stressed that NDU, corresponding to Hurka's V1 criterion, does not avoid what Michel and Pestieau (1998) called the Relative Repugnant Conclusion (see *supra*).

Moreover, Cowen (1989) and Broome (1992) emphasized that number-dampened utilitarianism suffers from the same weakness as average utilitarianism: it is not strongly separable between potential people. In other words, number-dampened utilitarianism, exactly as Hurka's "variable value view", makes the value of an additional individual depend on the number of already existing people, and on their welfare. Hence number-dampened utilitarianism suffers from the same problem as Hurka's criteria: no strong ethical argument seems to be able to justify that the social value of an additional person decreases with the size of the population, everything else being equal. Therefore NDU seems to lack strong ethical foundations: it seems to be an *ad hoc* solution without ethical support. Hence NDU, satisfying – relatively – our second requirement, does not satisfy our first requirement. It should be stressed that Ng is conscious that number-dampened utilitarianism lacks strong justifications: Ng personally considered NDU to be a compromise between what constitutes, according to

[60] Although Hurka's "variable value view" criteria V1 and V2 were discussed in section 7, I shall analyse NDU in greater depth here, because NDU constitutes a particular case (V1) of Hurka's "variable value view" criteria that has become quite popular in the literature.

[61] Given that Ng's number-dampened utilitarianism corresponds to Hurka's V1 criterion, differences between NDU and V2 correspond to the differences between V1 and V2.

[62] Under its "Absolute" formulation, but not under its "Relative" formulation (see *infra*).

[63] By "more compelling cases", Ng (1989a, p. 249) means "cases where the average utility of the added people is not very much lower than those of the pre-existing people, and the number of the pre-existing people has not become very large, so that most people find it very compelling to agree that the situation with the added people is better than the original situation." Such a remark is closed to what Hurka (1983) expressed regarding V1.

him, the "right moral principle", total utilitarianism (leading to the Repugnant Conclusion), and what he calls "our self-interests"[64].

9: Naverson's critique

As I stressed in the introduction, the difficulty of our inquiry into a "satisfactory" social welfare criterion depends on how we define a "satisfactory" criterion. So far, our inquiry has been guided by two requirements: strong ethical foundations and "reasonable" consequences. The assessment of the various social welfare criteria discussed so far has been carried out in the light of our intuition. Whatever our intuition constitutes a "good" guide or not for our inquiry (this is another difficult question), it should be stressed that our intuition might suggest us more precise ideas regarding the second requirement. Those more specific requirements were suggested by Naverson (1967) in his famous critique against utilitarian social criteria for optimal population choice. As we shall see, Naverson's critique holds against almost any social welfare criteria discussed in the previous sections.

In a nutshell, Naverson's critique can be summarized as follows. According to Naverson, utilitarianism is wrong when it considers that the addition of a happy person makes the society better off[65]. According to Naverson, an act is good only if it makes some people happier, and an act is bad only if it makes some people less happy. But in any case, what is crucial, according to Naverson, is that the existence of moral duties, in the utilitarian framework, depends on the existence of people. Hence, given that a person's own birth does not make that person happier (than before existing), it is false to believe that increasing the size of the population increases the society's happiness. Therefore, according to Naverson, an increase in the population cannot be valuable in itself[66]. That statement is justified by Naverson on the grounds that one cannot say that a state of affairs where one supplementary person exists is better than the same state of affairs, but without that person. Moreover, Naverson's statement is also rooted in another argument. As Naverson pointed out, a failure to do your moral duty of increasing the population does not affect anyone, because, in that case, the moral duty would "benefit" to someone who does not exist (yet). Thus if we assume that moral duties must be owed to people who exist, bringing up new people into existence can never be a moral duty.

However, Naverson also argued that we have a moral obligation not to bring into existence a child whose life would be miserable[67]. Indeed, in that case, our moral duty with respect to that miserable child would be to prevent him from having existed. Therefore, if we could predict that our child will be miserable if born, then our duty is not to have it.

Therefore, it follows from Naverson's critique that a "satisfactory" social welfare criterion should be compatible with the following asymmetry: it should regard the birth of a "happy" person as ethically neutral, but it should also regard the birth of a person whose life will be miserable as undesirable. Such an asymmetry, which seems to be intuitively

---

[64] Indeed, as discussed above, Ng (1989a) suggested that people could regard the Repugnant Conclusion as repugnant because of their inability to understand the implications of large numbers, or because of their "misplaced partiality". Therefore, according to Ng, who does not find the Repugnant Conclusion repugnant, the best moral principle would be total utilitarianism, and NDU would be a "compromise" between TU (morally desirable) and our own interests. However, as I argued in section 2, I personally think that the repugnancy of Parfit's Repugnant Conclusion might come from the problem asked (the Genesis Problem), or from our inability to escape our world, but I think that it cannot be reduced to the mere outcome of some "misplaced partiality".

[65] Naverson ignored the "indirect" effects due to someone's birth (on the parents' levels of "happiness"), but focused on the "direct" effects.

[66] On that issue, see Naverson (1967), pp. 63-69.

[67] On this issue, see Naverson (1967), pp. 69-71.

reasonable from an ethical point of view, is not easy to obtain. Indeed, as we shall see in the next section, social well-being criteria aimed at producing that asymmetry face deep difficulties. But before discussing those solutions to Naverson's critique, it might be worth emphasizing why the various utilitarian criteria analysed so far do not produce the required asymmetry[68]. The rest of this section is concerned with explaining why this is so. However, one should note that, in order to determine whether those social welfare criteria are subject to Naverson's critique or not, we must consider the problem of the addition of a person to an already existing population. Hence, we no longer deal here with the Genesis Problem, which assumed all people to be potential, but rather with the Actual Problem. In one sense, Naverson's critique, claiming that utilitarian criteria should care about existing people's welfare, and about their welfare only, is a critique that is more oriented against the Genesis Problem (assuming no already existing population) than against utilitarianism itself[69].

Regarding total utilitarianism, it should be stressed that the intuitively appealing asymmetry suggested by Naverson cannot be obtained by TU. Indeed, total utilitarianism considers the birth of a person with a happy life to be socially desirable, while, according to Naverson, it should be regarded as ethically neutral. Only an additional life with a negative utility would be regarded as undesirable. Hence total utilitarianism does not produce the desired asymmetry.

At first sight, average utilitarianism, which focuses only on the existing people's average welfare, seems to satisfy our intuitive asymmetry. However, as Broome (1996) noticed, average utilitarianism falls into the same trap as total utilitarianism: while there are two ways of increasing social welfare (given by the average level of utility), either by making existing people better off, or by adding people enjoying a utility level higher than the average, average utilitarianism considers – as opposed to Naverson's asymmetry – that the second way is acceptable: thus AU considers that an additional life is socially desirable, as long as the level of utility associated with that additional life is higher than the average utility. Therefore average utilitarianism violates our intuitive asymmetry requirement.

Regarding critical-level utilitarianism, it should be stressed that it also violates our intuitive asymmetry requirement. Indeed, as Broome (1996) underlined, CLU is neutral about adding a supplementary person only at the critical level of utility. However, if the utility of that additional person is higher (lower) than the critical level, then CLU regards the addition of that life as socially desirable (undesirable). One could argue that a high critical level would imply that no additional life is desirable in itself. But a high critical level would also imply that most additional lives would be socially undesirable. This is in conflict with what Naverson suggested: what is required is that the birth of a person with positive utility is ethically neutral, that is, neither socially desirable, nor socially undesirable. Obviously, critical-level utilitarianism, beyond all the criticisms mentioned in section 4, does not allow us to satisfy that reasonable requirement.

However, critical-band utilitarianism seems to be more appealing. Indeed, as underlined in section 5, critical-band utilitarianism generates an incomplete ranking of states of affairs. Therefore, CBU will consider that the addition of a supplementary person is neither desirable,

---

[68] On the ability of various social well-being criteria to produce our intuitive asymmetry, see Broome (1996).

[69] Hence it is not surprising that Naverson's critique is often interpreted as a critique of utilitarianism on the grounds of its impersonality. Instead of such an "impersonal" utilitarianism, Naverson recommended a "personal" utilitarianism ("person-affecting utilitarianism"), but such a "personal" utilitarianism can hold only if one assumes that there exists some population already (as opposed to what is assumed in the Genesis Problem). In other words, "personal" utilitarianism can hold only if one drops the Genesis Problem and starts thinking about the Actual Problem. But "personal" utilitarianism could also be criticized: as Glover (1990) underlined, a "personal" utilitarian, only concerned with existing people, would not matter whether there are people at all. Moreover, it might lead to contradictions (see section 10).

not undesirable as long as the utility level associated with that additional person lies within the interval [α^L, α^H], interval defined by the upper and lower critical levels of utility. Hence, CBU seems to satisfy our intuitive asymmetry requirement for additional lives whose utility levels lie in the interval. Provided the upper bound α^H of the critical interval is sufficiently high, CBU seems to satisfy our intuitive asymmetry requirement: only additional lives with very high utility levels would be regarded as socially desirable. If the lower bound α^L of the interval were zero, then only lives with negative utility levels would be regarded as socially undesirable. Therefore CBU, provided the bounds of the interval take particular values ($\alpha^L = 0$, α^H sufficiently high), allows us to be closed to the asymmetry suggested by Naverson. However, as Broome (1996) stressed, CBU allows us to obtain our required asymmetry only if we consider that "neutrality" means that the two states of affairs cannot be ranked against each other. However, if, by "neutrality", one means "equally good", then CBU does not allow us to obtain our asymmetry, because the fact that CBU cannot rank states of affairs does not necessarily imply that those states of affairs are equally good. Therefore adding a person with a utility level within the critical interval is "neutral" only if we consider that neutrality means "not ranked against". Broome (1996) considered that it is not clear what Blackorby, Bossert and Donaldson (1996) mean by "two states of affairs that cannot be ranked against each other". If this does not mean that those states are equally good, then one might expect from a social welfare criterion to rank them against each other, so that our asymmetry would then disappear. This made Broome talk about "cheating".

Regarding the two other criteria analysed above (Hurka's "variable value view" and Ng's number-dampened utilitarianism), it seems clear that they do not produce our required asymmetry. Hurka's V2 criterion might well assign declining weights to supplementary lives; those weights are still non-zero, so that the addition of a supplementary "happy" life is not ethically neutral. Ng's number-dampened utilitarianism also assigns a declining but positive value to additional "happy" lives, so that it violates our required asymmetry. It should be stressed that those two criteria, despite all their weaknesses[70], are not as far from our intuitive asymmetry as total utilitarianism is.

To summarize, almost all social welfare criteria discussed so far do violate our required asymmetry. Only critical-band utilitarianism seems to be preserved – at least to some extent – from Naverson's critique. Hurka and Ng's criteria might well seem to perform better than traditional criteria. However, they remain not fully satisfactory.

10: Solutions to Naverson's critique

At this stage, it might be worth discussing briefly some attempts to accommodate our intuitive asymmetry with a social welfare criterion. There is no place here to examine all the potential solutions aimed at accommodating Naverson's asymmetry with a social welfare criterion[71]. Therefore, I shall confine myself to discussing some of them[72].

A first attempt to express Naverson's asymmetry, proposed by Broome (1994, 1996), consists of what Broome calls the Constituency Principle[73]:

[70] See sections 7 and 8.

[71] Given that a social welfare criterion satisfying the asymmetry suggested by Naverson cannot help for solving the Genesis Problem (where it is assumed that no actual people exists), social welfare criteria discussed here deal with the Actual Problem.

[72] For a long excursion through various principles aimed at expressing Naverson's views, see Parfit (1984), pp. 394-417.

[73] Broome (1994), p. 229.

> "Suppose two states of affairs have the same population of people, except that an extra person exists in one who does not exist in the other. Suppose the extra person has a good life in the state in which she exists. Then one state is at least as good as the other if and only if it is at least as good for the people who exist in both."

Such a principle is very closed to Naverson's (1967) intuition: only the persons who exist in both states, forming a constituency, should count when one has to determine what is good or bad. However, as Broome (1994, 1996) showed, the Constituency Principle is false, because it leads to contradictions. More specifically, it contradicts the transitivity of the betterness relation. Let us have a look at Broome (1996)'s example 1[74]:

$W(a) = (1, 1, 2, \Omega, \Omega, \ldots)$
$W(b) = (1, 1, \Omega, \Omega, \Omega, \ldots)$
$W(c) = (1, 1, 1, \Omega, \Omega, \ldots)$

According to the Constituency Principle, a ~ b, because the two persons existing in states a and b enjoy the same utility level (equal to 1) under each state. Following the same rationale, we also conclude that b ~ c, because the two persons existing in states b and c enjoy the same utility level (equal to 1) under each state, while the utility of the extra person (the third one) under state c does not matter, because that person did not exist under state b. However, the Constituency Principle also suggests that a > c (because the third person existing in both states is definitely better off under state a than in state c, while the two other persons enjoy the same utility in each state), so that the transitivity of the betterness relation is violated. According to Broome, such a violation is not desirable, so that the Constituency Principle cannot be true. Therefore Broome (1994) examined three potential solutions that allow us to preserve the initial intuition provided by Naverson.

The first solution might be obtained by using an argument developed by Temkin (1982). Temkin simply argued that the betterness relation is intransitive. As Broome (1994) argued, such a view, if true, would support the Constituency Principle. Temkin's argument goes as follows. When two states of affairs are compared, this is always in the light of a particular criterion. However, when two other states of affairs are compared, this must not necessarily be in the light of the same criterion. Therefore, if one admits that different criteria are used for different comparisons, then the betterness relation might be intransitive.

However, Broome (1994) severely criticized Temkin's argument. Broome's counter-argument is that betterness is the comparative of goodness, so that betterness is necessarily transitive, because the comparative of any property must be transitive[75]. Moreover, Broome developed an example where the Constituency Principle leads to intransitive betterness and where the criteria used to compare the various states of affairs are exactly the same, so that Temkin's argument cannot serve as a defence of the Constituency Principle. Broome's example goes as follows[76]:

3A: $(w_1, w_2, \ldots, w_n, 1, 3, \Omega)$
3B: $(w_1, w_2, \ldots, w_n, 2, \Omega, \Omega)$
3C: $(w_1, w_2, \ldots, w_n, 3, \Omega, 1)$
3D: $(w_1, w_2, \ldots, w_n, \Omega, \Omega, 2)$
3E: $(w_1, w_2, \ldots, w_n, \Omega, 1, 3)$
3F: $(w_1, w_2, \ldots, w_n, \Omega, 2, \Omega)$

[74] See Broome (1996), p. 180, where Ω indicates that a person does not life in a particular history. Other examples might also be found in Broome (1996), p. 181, and in Broome (1994), pp. 230-232.
[75] According to Broome (1994), such a statement is true, whatever goodness is an "intrinsic" property or not.
[76] Broome (1994), example 3 p. 231.

According to the Constituency Principle, 3B > 3A, 3C > 3B, 3D > 3C, 3E > 3D, 3F > 3E, and 3A > 3F, so that the transitivity of the betterness relation is violated in this example. However, as Broome argued, Temkin's argument of intransitivity of the betterness relation cannot be used here. Indeed, states 3A and 3B are compared by looking at the welfare of the person n+1, and states 3B and 3C are also compared by looking at the welfare of person n+1, so that, in each of those two comparisons, the criterion that is used is the same. Hence, Temkin's intransitivity argument – based on the use of distinct criteria for different comparisons – cannot hold here: transitivity should prevail. Transitivity would suggest that 3C > 3A. However, Broome argued that 3A and 3C should be equally good, because only the identities of the persons who exist and the distribution of welfare between the existing persons are different, but total welfare is the same. Hence the Constituency Principle, suggesting by transitivity that 3C > 3A, contradicts Broome's claim that 3C ~ 3A. According to Broome, that contradiction casts some doubts on the adequacy of the Constituency Principle, and Temkin's argument cannot intervene here in order to "save" the Constituency Principle. In my view, Broome's claim that "3A and 3C should be regarded as equally good because they generate the same total welfare" can be regarded as being based on exactly what Naverson wanted to reject: total welfare considerations completely independent of which person is currently alive or not. Hence this might not be very surprising that the introduction of the claim that "3A and 3C should be equally good" directly leads to a contradiction with what the Constituency Principle recommends. Therefore, I personally think that Broome's example, although it shows that Temkin's argument cannot be used to "save" the Constituency Principle, might not show that the Constituency Principle is self-contradictory, because it only reaches a contradiction if some total welfare consideration is introduced. Hence one could look at Broome's example as an illustration of the contradicting recommendations of the Constituency Principle and total utilitarianism.

The second solution discussed in Broome (1994) consists of saying that the betterness relation might be conditional in a particular sense. Indeed, when we think about the asymmetry suggested by Naverson, it might be reformulated in terms of "conditional goodness", that is, in terms of goodness conditional on existence: bringing a "happy" person to existence is ethically neutral, but bringing a "unhappy" person is not ethically desirable (provided we add a person, that person must have a good life). However, Broome (1994) rejected such an interpretation, on the grounds that the idea of conditional good is incoherent. Moreover, Broome (1994) developed an example where conditional betterness leads to intransitivity in the strict betterness relation, which is not acceptable[77].

A third solution discussed by Broome (1994) is the one based on Dasgupta's (1988, 1993) population-relative ethics. As Broome (1996) underlined, Dasgupta (1994) does not adhere to Naverson's asymmetry. However, Broome considered that Dasgupta's population ethical theory could be used in order to obtain our intuitively desirable asymmetry. In a nutshell, Dasgupta's population-relative ethics can be summarized as follows. According to Dasgupta, the moral basis for Different Number Choices is generation-relative. There is no unique betterness relation: the moral betterness relation is relative to a particular population, and it might vary according to the point of view[78]. Moreover, Dasgupta (1988) stressed that the Genesis Problem, assuming all people to be potential, is not the right problem to solve. Rather, the relevant population problem to study – the Actual Problem – consists of choosing an optimum population size given that there are actual people already existing (the current population size is taken as given). In such a problem, the current population deliberates about future population sizes and future welfare. In order to determine the optimal population sizes,

[77] See Broome (1994), p. 238.

[78] Such an agent-relative (or in this case, generation-relative) approach to ethics, which differs completely from the other approaches discussed above, was previously suggested by Sen (1982, 1983).

Dasgupta's theory recommends a two-stage procedure[79]. In the first stage, all the potential options would be classified in sets according to their population. Then, one should select the "best" option in each of the sets, that is, the one that maximizes the relative value function of that set's population (thus the one that maximizes the unweighted total welfare of that particular population). Then the second stage consists of the selection, by the decision-making population, of the "best" option among the pre-selected options (that "best" option being the one that maximizes the population's own relative value function[80]). As Broome (1996) noticed, such a procedure might generate our intuitive asymmetry, but only in the particular case where the decision-making population assigns a zero weight to the welfare of any people outside its population[81].

However, Broome (1994, 1996) argued against Dasgupta's approach that, although there might be some justification for a population to give more weight to its members than to other people, there is no justification for assigning a zero weight to any other people[82]. According to Broome, such an extreme position is not defendable. As Broome rightly pointed out, Naverson's intuition is that a person who does not exist has no interest to be brought into life (because being brought to life would not make him better off), but not that such a person has an interest that is ignored. Therefore Broome claimed that an adequate population theory cannot be based on community-relativity. Moreover, Broome criticized Dasgupta's two-stage procedure, on the grounds that such a procedure does not imply that each population ought to do the best it can according to its own betterness ordering[83]. Therefore, according to Broome, Dasgupta's relativist theory, although it allows us to obtain the desired asymmetry, seems to need stronger justifications.

To summarize, none of the "solutions" discussed here seems to allow us to accommodate our intuitively appealing asymmetry with a social welfare criterion for optimal population choice. Neither the use of Temkin's argument (to defend the Constituency Principle), nor the interpretation in terms of conditional betterness (to avoid contradictions) is fully convincing. Moreover, the use of Dasgupta's generation-relative ethics in this context might not be completely satisfactory. Therefore, it follows from this section that if we require our social welfare criterion to satisfy the asymmetry suggested by Naverson (1967), finding a "satisfactory" social welfare criterion for population choices (Actual Problem) seems difficult. Hence one could wonder whether the requirement of our intuitively appealing asymmetry is really reasonable. Maybe our intuition might be misleading or simply too demanding. Or maybe it is the requirement of ethical foundations that might be too demanding.

11: Conclusions

[79] One can find a complete discussion of that two-stage procedure in Broome (1994, 1996).

[80] The decision-making population's own relative value function consists of the value function (the weighted sum of the well-being of all the people who exist) where the decision-making population assigns itself a higher weight than to the other populations.

[81] On Dasgupta's generation-relative ethics, see Dasgupta (1993), chapter 13. Dasgupta underlined that one cannot talk about incoherence regarding successive population choices made within his framework. Indeed, there is no incoherence (nor intransitivity), because the population making the choices of future population sizes changes at each period, so that the moral point of view changes each time.

[82] Against that attack, Dasgupta (1993) argued that each generation, one at a time, successively "plays God", in choosing the next generation's size, so that no generation is privileged. Moreover, Dasgupta (2001) underlined that it is not necessary that each generation assigns a zero weight to other (potential) people. Indeed, one should remind that such an extreme position is only required in order to allow us to obtain our "desired" asymmetry (but this is not Dasgupta's initial goal).

[83] According to Broome, what the population ought to do, according to the two-stage procedure, is determined by another population's betterness relation rather than its own ordering.

There is no place here to summarize our long and difficult inquiry into a "satisfactory" social welfare criterion for optimal population choices. However, several points should be stressed.

Firstly, it should be stressed that "advances" made in the literature of utilitarian population ethics are of great theoretical interest not only for utilitarianism itself, but also for any consequentialist or consequences-based ethical theory[84]. Indeed, the issue at stake here was the one of finding an adequate combinatory structure in order to deal with Different Number Choices, independently of any consideration regarding the informational basis used (utility). Therefore the literature to which this survey was dedicated asks questions whose importance and relevancy extend far beyond the realm of utilitarianism. In other words, any ethical theory based on consequences would gain from "advances" in utilitarian population ethics, because the difficulties faced by utilitarianism to guide Different Number Choices would also be faced by other consequences-based ethical theories.

Secondly, our discussions in sections 2 and 3 might suggest that neither total utilitarianism (facing Parfit's Repugnant Conclusion), nor average utilitarianism (facing the Mere Addition Paradox and having controversial ethical foundations) seem to satisfy our two requirements of strong ethical foundations and "reasonable" consequences. Moreover, it was argued in section 4 that critical-level utilitarianism does not seem to be more successful (no strong foundations and some counter-intuitive implications), while critical-band utilitarianism seems to provide better consequences, but without satisfying our first requirement (section 5).

Thirdly, it was showed in section 6, in the light of Ng's contribution, that Parfit's four "basic" requirements might be very demanding, so that there seems to exist a quasi-impossibility of finding a "Theory X". This result is fundamental, because it shows that the combination of a limited number of basic requirements can lead us to an impossibility to find a satisfactory social welfare criterion, even if we completely neglect our first requirement (strong ethical foundations). Hence one should maybe not be too demanding regarding what one can expect from a social welfare criterion.

Fourthly, some candidates to "Theory X" were discussed in sections 7 and 8. It was argued that those criteria, despite their relatively good consequences, lack strong ethical foundations: they seem to be *ad hoc*. This might suggest a second conclusion: there might exist a "tension" between our two requirements (strong ethical foundations and "reasonable" consequences), so that an improvement on the grounds of the second criterion might be achieved only at the cost of some loss of ethical foundations or meaning, so that the most adequate criteria might be the ones that look the most arbitrary.

Fifthly, it was stressed in section 9 that Naverson's critique weakens almost any of the criteria discussed in the previous sections (except, to some extent, critical-band utilitarianism, and, to another extent, number-dampened utilitarianism and Hurka's V2 criterion). Then, section 10 showed the difficulties to construct a criterion based on the asymmetry suggested by Naverson, and thus suggested another conclusion of this inquiry: taking the asymmetry requirement as a part of the second criterion – and thus turning to the Actual Problem – does not seem to make the inquiry into a "satisfactory" criterion easier.

It follows from this that the present survey might suggest three main conclusions: firstly, it might be difficult to find a SWF generating "reasonable" consequences; secondly, such a SWF might become even more difficult to find if one adds the requirement of exhibiting strong ethical foundations; thirdly, when one turns to the Actual Population Problem, finding

[84] I use the word "advances", because I personally think that important results have been derived in that literature. My statement seems to take for granted that "advances" or "progress" in population ethics can be made. However, the question of whether any "advances" or "progress" could be made in ethics is a difficult question whose answer would require further investigations into the nature of ethical knowledge. That difficult question is widely discussed in Mackie (1977) and Williams (1985).

a SWF satisfying Naverson's asymmetry requirement might not be an easy task. What attitude should we have in front of such conclusions? Here I shall consider briefly two attitudes: on the one hand, a questioning of the adequacy of our requirements, and, on the other hand, a questioning of the adequacy of utilitarianism as an ethical theory of public choices.

On the one hand, it could be argued that the conclusions of this survey might result from the definition and the number of the requirements in the light of which SWF were assessed: the more numerous and the more "strict" the requirements are, the more difficult – if not impossible – the inquiry into a "satisfactory" SWF is. More precisely, a weak version of that position towards our result consists of saying that our difficulties to find a "satisfactory" criterion might be due to the fact that our requirements are quite demanding. A stronger version of that position suggests that it might also be the case that some requirements contradict each other, leading automatically to some "impossibility" results. That possibility might emerge from the fact that our requirements – and especially the second one – were mainly based on "intuitions" or "ethical beliefs"[85]. Those "intuitions", aimed at evaluating various utilitarian criteria, or, in other words, aimed at overcoming what Mackie (1977) called an "indeterminacy" within utilitarianism, had to be independent on any utilitarian consideration. However, in such a difficult issue as population ethics, this might not be surprising that our intuitions might sometimes be wrong, or misleading[86]. The possibility that some of our intuitions are wrong or contradict each other might explain why we cannot find a criterion satisfying all our requirements[87]. Therefore one explanation for our difficulties to find a "satisfactory" SWF might come from our intuitive requirements themselves, whatever we consider that those contradict themselves (stronger version) or not. A solution to our difficulties could be to try to discover what lies behind our intuitions, in order to identify the underlying contradictions if those exist, but this does not constitute an easy task, so that the only "feasible" way to find a "satisfactory" utilitarian criterion would be to reduce our ambitions, and to give up some (maybe contradicting) requirements. But which requirements should be given up? One could also be more sceptical towards our whole inquiry, and adopt what I would call the "sceptical" version of the first attitude. Indeed, one could wonder whether it could be possible to discriminate between various SWF by means of – even a restricted set of – "objective" criteria. One could argue, as Mackie (1977), that it will always remain impossible to discriminate between various SWF on "objective" grounds, so that the only thing a utilitarian could do is simply to choose, in an arbitrary way, one combinatory structure to deal with population choices. In Mackie's (1977) words[88]:

> "All these difficulties and indeterminacies [including the treatment of Different Number Choices] tell, in the first place, only against the claim that utilitarianism offers a peculiarly unitary and systematic basis for morality. A utilitarian can simply decide which of the various options to take up, and he can plausibly argue that rival views are subject to similar indeterminacies".

Hence, according to Mackie, there might exist no relatively "objective" way to overcome the "indeterminacy" in utilitarianism that was at stake in this survey. Mackie's view is quite

[85] Indeed, our assessments of the various criteria were mainly based on our "intuitions", which, for instance, "told" us which consequences were ethically appealing or not.

[86] As Williams (1985) underlined, the role of intuition in ethics is far from being uncontroversial. For a critique on the role of intuition in ethics, see Toulmin (1950).

[87] For instance, the intuitions suggesting that overpopulation is "bad" (Repugnant Conclusion) and that underpopulation is "bad" (the Mere Addition Paradox) might not only seem to be arbitrary, but also – at least to some extent – contradictory, so that no utilitarian guide to optimal population choices satisfying those two requirements could be found. However, there seems to be no consensus on the extent to which those intuitions are contradictory: obviously, Parfit (1984) and Ng (1989a) disagree on that point (see *supra*).

[88] Mackie (1977), p. 129.

pessimistic regarding the possible justification of any utilitarian criterion[89]. But does there exist no way to overcome, on "objective" grounds, that "indeterminacy"? That question, which raises the difficult issue of the status of ethical knowledge and of its possible or impossible "objectivity", goes far beyond the scope of this survey, so that there is no place here to analyse it in depth. However, a less ambitious question would be to wonder whether, in the light of our own inquiry, adopting a pessimistic view seems justified. In my view, adopting a pessimistic or sceptical conclusion does not seem to be fully justified in the light of the present survey. What our inquiry illustrates is not that no kind of relatively "objective" discrimination between various SWF is possible, but rather that such a discrimination is only possible provided our "intuitive" requirements are more clearly defined and provided those are – at least to some extent – restricted. Therefore there might be no reason for adopting an extremely pessimistic or sceptical point of view.

On the other hand, one could interpret our difficulties to find a "satisfactory" SWF in a different way. Rather than questioning the adequacy of the requirements we used, one might simply argue that what the present survey suggests is that utilitarianism – whatever its formulation – might not necessarily be able to provide "satisfactory" answers to all thinkable questions, and especially, to such difficult questions as optimal population issues. According to that interpretation, the difficulties faced in this survey would simply remind us that no ethical theory – including utilitarianism – can answer all ethical questions in a fully satisfactory way. Indeed, as Professor Mirrlees (1982) wrote[90]:

> "The utilitarian method does not answer all questions. Is there any reason to think we are in a better position to decide how much to spend on kidney machines, than we are to decide how long this universe will last? That one does not know the answer to many moral questions is a reason for developing systematic procedures. But even after attempted analysis, not knowing may be the correct answer. Two examples of this are the treatment of handicapped people and the question of optimum population."

Hence utilitarianism, as any human construction, inevitably exhibits the limits and weaknesses of its human creators. But those limits should not be regarded as something specific to utilitarianism[91]. As I stressed above, any consequentialist or consequences-based ethical theory – even outside the welfarist domain – would face the same difficulties as the ones faced by utilitarianism in front of what Parfit (1984) called Different Number Choices. Therefore the limits of utilitarianism in the field of population ethics cannot, in my view, be regarded as an impasse of utilitarianism itself, but rather as the limits of consequences-based theories in general in front of Different Number Choices.

[89] This view suggests that one should not base an ethical theory on consequences, but rather, for instance, on rights, as Mackie argued.
[90] Mirrlees (1982), p. 80.
[91] Mackie (1977) also seemed to acknowledge that fact, when he claimed, in the quotation above; "A utilitarian can simply decide which of the various options to take up, and he can *plausibly* argue that rival views are subject to similar indeterminacies" (Mackie, 1977, p. 129, italics added by the author).

13: Bibliography

Barry, B. (1989): "Theories of Justice", University of California Press, vol. 1.

Bentham, J. (1789): "An introduction to the Principles of Morals and Legislation", in Mill, J.S.: "Utilitarianism", edited by M. Warnock (1962), Fontana Press.

Birnbacher, D. (1988): "Verantwortung für Zukünftige Generationen", Stuttgart, (french translation: "La responsabilité envers les générations futures", 1994, Presses Universitaires de France, Paris).

Blackorby, C. & Donaldson, D. (1984): "Social Criteria for Evaluating Population Change", *Journal of Public Economics*, 25, pp. 13-33.

Blackorby, C., Bossert, W. & Donaldson, D. (1995): "Intertemporal Population Ethics: Critical-level Utilitarian Principles", *Econometrica*, 63, pp. 1303-1320.

Blackorby, C., Bossert, W. & Donaldson, D. (1996): "Quasi-orderings and Population Ethics", *Social Choice and Welfare*, 13, pp. 129-150.

Broome, J. (1991): "Weighing Goods", Basil Blackwell.

Broome, J. (1992): "Counting the Cost of Global Warming", London, White Horse Press.

Broome, J. (1994): "The Value of a Person", *Proceedings of the Aristotelian Society*, supplementary volume 68, pp. 167-185, republished in Broome, J. (1999) "Ethics out of Economics", Cambridge University Press, chapter 15.

Broome, J. (1996): "The Welfare Economics of Population", *Oxford Economic Papers*, 48, pp. 177-193.

Broome, J. (1999): "Ethics out of Economics", Cambridge University Press.

Cowen, T. (1989): "Normative Population Theory", *Social Choice and Welfare*, 6, pp. 33-43.

Dasgupta, P.S. & Heal, G.M. (1979): "Economic Theory and Exhaustible Resources", Cambridge, Cambridge University Press.

Dasgupta, P.S. (1988): "Lives and Well-Being", *Social Choice and Welfare*, 5, pp. 103-126.

Dasgupta, P.S. (1993): "An Inquiry into Well-being and Destitution", Oxford University Press.

Dasgupta, P.S. (1994): "Savings and Fertility: Ethical Issues", *Philosophy and Public Affairs*, pp. 99-127.

Dasgupta, P.S. (2001): "Human Well-being and the Natural Environment", Oxford, Oxford University Press.

Diamond, P.A. (1967): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility: Comment", *Journal of Political Economy*, 75, pp. 765-766.

Glover, J. (1990) (eds.): "Utilitarianism and its Critics", London, MacMillan.

Gottlieb, M. (1945): "The Theory of Optimum Population for a Closed Economy", *Journal of Political Economy*, 53, pp. 289-316.

Hammond, P.J. (1988): "Consequentialist Demographic Norms and Parenting Rights", *Social Choice and Welfare*, 5, pp. 127-145.

Harsanyi, J.C. (1953): "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking", *Journal of Political Economy*, 61, pp. 434-435.

Harsanyi, J.C. (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", *Journal of Political Economy*, 63, pp. 309-321.

Hurka, T. (1983): “Value and Population Size”, *Ethics*, 93, April, pp. 496-507.

Kolm, S-C. (1969): “The Optimal Production of Social Justice”, in Margolis, J. & Guitton, H. (eds.), “Public Economics”, London, MacMillan.

Mackie, J.L. (1977): “Ethics: Inventing Right and Wrong“, Penguin Books.

Michel, P. & Pestieau, P. (1998): “Optimal Population Without Repugnant Aspects”, *Genus*, LIV, n°3-4, pp. 25-34.

Mill, J.S. (1859): “On Liberty”, in J.S. Mill: “Utilitarianism”, edited by M. Warnock (1962), Fontana Press.

Mirrlees, J. (1982): “The Economic Uses of Utilitarianism”, in Sen, A.K. & Williams, B: “Utilitarianism and beyond” (1982), Cambridge University Press, pp. 63-84.

Naverson, J. (1967): “Utilitarianism and New Generations”, *Mind*, 76, pp. 62-72.

Ng, Y-K. (1986): “Social Criteria for Evaluating Population Change: an Alternative to the Blackorby-Donaldson Criterion”, *Journal of Public Economics*, 29, pp. 375-381.

Ng, Y-K. (1989a): “What should we do about Future Generations? Impossibility of Parfit’s Theory X” *Economics and Philosophy*, 5, pp. 235-253.

Ng, Y-K. (1989b): “Hurka’s Gamble and Methuselah’s Paradox: a Response to Cowen on Normative Population Theory”, *Social Choice and Welfare*, 6, pp. 45-49.

Parfit, D. (1982): “Future Generations, Further Problems”, *Philosophy and Public Affairs*, 11, pp. 113-172.

Parfit, D. (1984): “Reasons and Persons”, Oxford, Oxford University Press.

Ponthiere, G. (2003): “Should We Discount Future Generations’ Welfare? A Survey on the ‘Pure’ Discount Rate Debate”, *CREPP Working Paper*, n°--, University of Liege.

Rawls, J. (1971): “A Theory of Justice”, Cambridge, Massachusetts, Harvard University Press.

Sen, A.K. (1970): “Collective Choice and Social Welfare”, San Francisco, Holden Day.

Sen, A.K. (1982): “Rights and Agency”, *Philosophy and Public Affairs*, 11, pp. 3-38.

Sen, A.K. (1983): “Evaluator Relativity and Consequential Evaluation”, *Philosophy and Public Affairs*, 12, pp. 111-132.

Sen, A.K. (1984): “Well-being, Agency and Freedom: the Dewey Lectures 1984”, *Journal of Philosophy*, volume LXXXII, n° 4, April 1985, pp. 169-221.

Sen, A.K. (1986): “Social Choice Theory”, in Arrow, K.J. & Intriligator, M.D. (eds.) (1986) “Handbook of Mathematical Economics: volume III”, Amsterdam, North-Holland, pp. 1073-1181.

Sen, A.K. (1987): “On Ethics and Economics”, Basil Blackwell.

Sidgwick, H. (1874): “The Methods of Ethics”, London, MacMillan.

Temkin, L.S. (1982): “Intransitivity and the Mere Addition Paradox”, *Philosophy and Public Affairs*, 11, n°2, pp. 113-172.

Temkin, L.S. (1992): “Intergenerational Inequality”, in Laslett, P. & Fishkin, J.S. (eds), “Philosophy, Politics and Society: Sixth Series: Justice between Age groups and Generations”, Yale University Press, 1992.

Toulmin, S.E. (1950): “An Examination of the Place of Reason in Ethics”, Cambridge University Press.

Weymark, J.A. (1991): “A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism”, in Elster, J. & Roemer, J. (eds.) (1991) “Interpersonal Comparisons of Well-Being”, Cambridge University Press.

Williams, B. (1985): “Ethics and the Limits of Philosophy”, Fontana Press.