Laboratoire REGARDS (EA 6292)
Université de Reims Champagne-Ardenne

Working paper n° 1-2014

Utilitarianism, Prioritarianism and Nonlinear Social Welfare Functions: Should We Accept Bernoulli's Hypothesis?

Cyril Hédoin*

* Maitre de conférences en sciences économiques, REGARDS (Université de Reims Champagne-Ardenne)

Abstract
Utilitarianism and prioritarianism are two of the most important consequentialist theories in ethics. Prioritarians hold that, everything else equal, it is better that benefits go to the worse-off people than to the better-off, *i.e.* a given amount of good counts more the less well-off are the people it goes to. However, axiomatic defenses of utilitarianism relying on expected utility theory, such as those of John Harsanyi (1955) and John Broome (1991), offer a surprising and counterintuitive objection to prioritarianism: the distinction between an *amount* of good and how this amount *count* is empty and meaningless. This objection ultimately depends on the validity of "Bernoulli's hypothesis" which states that an alternative is better for someone than another one if its expected goodness is higher. This article weights the reasons for and against Bernoulli's hypothesis. And concludes that expected utility theory does not entail utilitarianism. This seems to be a particular instance of the more general rule that ethics does not supervene on rationality.

Mots clés : Utilitarianism – Prioritarianism – Expected Utility Theory – Bernoulli's hypothesis – Strong independence – John Broome

Les working papers d'économie et gestion du laboratoire Regards sont édités après présentation en séminaire et validation par deux relecteurs internes, sous la responsabilité du conseil de laboratoire.

Laboratoire d'Economie et
Gestion REGARDS (EA 6292)
Université de Reims
Champagne-Ardenne
UFR de sciences économiques,
sociales et de gestion
57B Rue Pierre Taittinger
51096 Reims

Directeur : **Martino Nieddu**

Responsable de l'édition des *working papers* **: Cyril Hedoin**

WP 1-2014

Utilitarianism, Prioritarianism and Nonlinear Social Welfare Functions: Should We Accept Bernoulli's Hypothesis?

Cyril Hédoin[*]

Do not quote without permission

Abstract: Utilitarianism and prioritarianism are two of the most important consequentialist theories in ethics. Prioritarians hold that, everything else equal, it is better that benefits go to the worse-off people than to the better-off, *i.e.* a given amount of good counts more the less well-off are the people it goes to. However, axiomatic defenses of utilitarianism relying on expected utility theory, such as those of John Harsanyi (1955) and John Broome (1991), offer a surprising and counterintuitive objection to prioritarianism: the distinction between an *amount* of good and how this amount *count* is empty and meaningless. This objection ultimately depends on the validity of "Bernoulli's hypothesis" which states that an alternative is better for someone than another one if its expected goodness is higher. This article weights the reasons for and against Bernoulli's hypothesis. I show that the case for Bernoulli's hypothesis depends on a particular formal conception of goodness which is not mathematically supported by expected utility theory. It follows that the latter does not entail utilitarianism. This seems to be a particular instance of the more general rule that ethics does not supervene on rationality.

Keywords: Utilitarianism – Prioritarianism – Expected Utility Theory – Bernoulli's hypothesis – Strong independence – John Broome

Word count: 11759

[*] Associate professor of economics, REGARDS research center, University of Reims Champagne-Ardenne, France.
cyril.hedoin@univ-reims.fr

Utilitarianism, Prioritarianism and Nonlinear Social Welfare Functions: Should We Accept Bernoulli’s Hypothesis?

1. Introduction

In his seminal article “Equality or Priority?”, Derek Parfit (1991) draws a clear distinction between various forms of egalitarianism and what he calls the “Priority View”. Parfit summarizes the latter with the following statement: “Benefitting people matters more the worse off these people are” (Parfit 1991, 101). Parfit’s Priority View gave rise to what is now known as *prioritarianism*. Though some authors continue to hold that the distinction between equality and priority is misguided and that prioritarianism is not qualitatively different than egalitarianism, it is generally agreed that to distinguish between these two views is useful and relevant. *Prima facie*, the difference between them is not transparent. As Parfit himself notes, egalitarians obviously agree with the statement that it is better to benefit people the worse off they are. But prioritarians and egalitarians do not make this claim for the same reasons. While egalitarians value equality *per se* and value benefiting the worse off as a way to reduce inequality, prioritarians are not interested in reducing inequality in itself. The badness associated to the fact that some people are worse off than others do not stem from the resulting inequality, but rather from the fact that they are worse off than they might have been (Parfit 1991, 104). In other words, the priority view is only interested in people’s *absolute* level of wellbeing, while egalitarians care for the *relative* level of each person’s wellbeing.

In this article, I will take for granted that the above distinction is relevant. I will also take for granted that prioritarianism pertains to the realm of consequentialist theories of morality. As a result, the Priority View must be amenable to a formalization through a *theory of the good*: other things equal, a state of affairs is better the more benefits the worse off people receive. Conceived as a theory of the good, prioritarianism is then a natural alternative to utilitarianism. Utilitarianism is neutral toward the distribution of the good. On the other hand, all things equal, prioritarianism favors a more egalitarian distribution. In his book *Weighing Goods*, John Broome (1991) derives a utilitarian formula from a variant of John Harsanyi’s aggregation theorem (Harsanyi 1955) stating that an alternative A is better than an alternative B if the sum of individual goods is superior in A than in B. Crucially, Broome argues that the “Utilitarian distribution principle”, as he calls it, is inimical to any form of “communal egalitarianism”, including prioritarianism. More importantly, Broome’s axiomatic approach leads him to conclude that prioritarianism is incoherent if one is ready to accept the various assumptions leading to the utilitarian principle. Broome’s disturbing conclusion follows in particular from a critical assumption, *Bernoulli’s hypothesis*. It can be understood as a statement about the nature of the good which indicates that the individual betterness relation between two alternatives depends on the expected goodness of those alternatives. As Broome explicitly recognizes, Harsanyi’s aggregation theorem has an utilitarian meaning only if Bernoulli’s hypothesis holds.

My purpose in this article is to show that Bernoulli's hypothesis seems hard to sustain for very technical reasons that has been informally presented by Sen (1986) and more fully elaborated by Weymark [(1991); (2005)]. The consequence is that it is no longer a mathematical necessity for the social welfare function representing general goodness to be linear in individuals' goodness. Since prioritarianism exhibits a nonlinear combination of persons' good, this tends to vindicate the Priority View as a credible alternative to utilitarianism, as far as consequentialism is concerned. More precisely, I argue that what I call the Broome-Harsanyi argument for the Bernoulli's hypothesis only weakly support utilitarianism, unless one is ready to defend an extreme form of it, namely "complete utilitarianism". Since complete utilitarianism relies on a debatable (but not necessarily unsustainable) metaphysical theory of personal identity, this lends support for prioritarianism. Overall, the case for Bernoulli's hypothesis depends on a particular formal conception of goodness which is not mathematically supported by expected utility theory. It follows that the latter does not entail utilitarianism.

The rest of the article is made of eight sections. I distinguish between two kinds of prioritarianism in the next section: "individualistic prioritarianism" and "communal prioritarianism". The latter results from the rejection of another of Broome's axioms, the principle of personal good (Rabinowicz 2002). I make it clear that my discussion applies solely to individualistic prioritarianism, even though communal prioritarianism may be an attractive alternative. The third section presents Broome's axiomatic account of the utilitarian principle and explains why it refutes (individualistic) prioritarianism. The fourth section briefly considers a possible but ultimately inadequate defense of Bernoulli's hypothesis according to which this hypothesis follows from the quantitative nature of goodness. The fifth section presents the Broome-Harsanyi argument for Bernoulli's hypothesis. I explain in the sixth section why this argument is inadequate at the face of Sen's and Weymark's technical remarks regarding cardinal utility functions. I suggest in the seventh section that to keep Bernoulli's hypothesis, one should probably be ready to defend complete utilitarianism against the priority view. The eighth section shows that rejecting Bernoulli's hypothesis may be sometimes the only way to make individualistic prioritarianism compatible with the strong independence axiom of expected utility theory. I briefly conclude in the final section.

2. "Individualistic" Prioritarianism and "Communal" Prioritarianism

As a theory of the good, prioritarianism can be identified to the following statement: a given *amount* of good *counts* the more as it is attributed to people with an absolutely low level of wellbeing. I emphasize "amount" and "count" in the preceding sentence because this is what separates prioritarianism from utilitarianism: other things equal, it is better to increase the level of wellbeing of worse off people, while for utilitarianism distributive issues do not matter. An easy and straightforward formalization of prioritarianism is given by the following social welfare function:

(1) $g(x) = \sum_i w(g_i(x))$

Equation (1) says that the general or overall goodness g of an alternative x is the sum of the increasing transforms w of individual goodness g_i with $i = 1, \ldots, n$ the members of the relevant population. In this paper, I generally treat any alternative x as a *prospect* which attributes an outcome o_j for each possible state of nature s_j. Thus, an alternative x has the form (o_1, p_1; o_2, p_2; …; o_m, p_m) for m possible outcomes and states of the world, with p_j the probability attached to state of nature s_j.[1] If the function w is strictly concave, then the contribution of individual goodness to the overall goodness is marginally decreasing. This leads to the priority view since equation (1) indicates that any given increase in individual goodness of person i the more contributes to overall goodness the lower i's absolute level of individual goodness is. More formally, all forms of prioritarianism corresponding to equation (1) imply that Pigou-Dalton transfers of individual goodness are always improvements (McCarthy 2006, 340).[2]

There are at least two ways to arrive at (1). The first one is proposed by Rabinowicz (2002) who argues for the rejection of one of the key assumption of Broome's utilitarian distribution principle, the principle of personal good. This principle indicates that overall goodness supervenes on individual goodness in the following way:

a) Two alternatives A and B that are equally good for each individual are equally good overall.
b) If an alternative A is strictly better than an alternative B for at least one individual i, and if A is at least as good as B for every other individuals j, then A is better than B overall.

Broome [(1991); (2004)] distinguishes between two versions of the principle of the personal good. The weaker version only holds for *outcomes* or, equivalently, for sure prospects. Sure prospects attribute the same outcome o_j* to every state of nature such that the alternative x leads to the outcome o_j* with certainty. The stronger version applies for *risky prospects*. As Broome recognizes, this version is more contentious. In particular, it requires that people agree on the underlying probabilities or, more convincingly, that probabilities have an objective meaning such that they can be held constant for any prospect.[3] Rabinowicz argues that prioritarians should accept the weaker version of the principle of personal good but reject it for prospects. This move has several important implications. A first direct consequence is that it is now possible that for two prospects x and y and two individuals i and j, x is strictly

[1] Of course, probabilities must sum up to 1, *i.e.* $\sum_j p_j = 1$.

[2] A Pigou-Dalton transfer of individual goodness leaves everyone's situation in terms of goodness unchanged except for two persons i and j. It makes the better off person i worse off by transferring a given amount of goodness to the worse off person j, while leaving i relatively better off than j. Thus, Pigou-Dalton transfers do not affect the sum of individual goodness but change the distribution of goodness across people toward more equality.

[3] Contrary to Broome's, Harsanyi's aggregation theorem relies on the Pareto principle where probabilities are subjective beliefs. However, if two persons do not have consistent subjective beliefs, it is then possible that they both prefer a prospect A over a prospect B but that actually A is better than B only for one of the two persons. More generally, for risky prospects, the Pareto principle is inconsistent with the assumption that persons have coherent preference over alternatives (see Broome 1991, 152-3).

better than *y* for both persons in terms of individual goodness, but that regarding general goodness *y* is better than *x*. Consider the following two prospects:[4]

Figure 1

	x			*Y*	
	s_1	s_2		s_1	s_2
i	9	11		16	5
j	11	9		5	16

We assume that states are equiprobable (*i.e.* $p_1 = p_2 = ½$). The numbers in the matrix indicate the individual goodness for each outcome and each person. Rabinowicz argues that, as far as individual goodness is concerned, the betterness relation between prospects should depend on the prospects' expected goodness.[5] Therefore, prospects *y* is better than *x* for both persons *i* and *j*. However, regarding overall goodness, Rabinowicz suggests that it is quite sensible to consider that *x* is better than *y*, *in spite of the fact that x's expected goodness is lower than j's*. An egalitarian could also make this claim on the ground that whatever the actual state of nature is, prospect *x* guarantees a more equal outcome than *y*. But a prioritarian should argue otherwise: prospect *x* guarantees that the worst off person is largely better off than she would be in the outcomes attached to prospect *y*. More precisely, for each possible outcome, the partial transfer of goodness from the better off person in prospect *y* to the worse off person in prospect *x* is an improvement because the leveling up of the worst off's situation *more than counterbalances* the small diminution of the sum of individual goodness. Obviously, this conclusion implies that the principle of personal good does not hold for prospects.

The rejection of the principle of personal good leads to what I call "communal prioritarianism". I will contrast communal prioritarianism with "individual prioritarianism". The latter, contrary to the former, accepts the strong version of the principle of personal good but rejects Bernoulli's hypothesis. What is the point of considering individual prioritarianism rather than its communal version? The main motivation lies in the fact that Broome's argument that prioritarianism (or "additive egalitarianism" as he calls it in *Weighing Goods*) is incoherent takes the strong principle of personal good for granted. By the way, this is what distinguishes prioritarianism from egalitarianism according to Broome. Moreover, I think that the principle of personal good has an intuitive appeal, even when applied to prospects. Broome (1991, 167-173) outlines an argument vindicating this intuition. All in all, I think that while the principle of personal good can be argued for and against, Bernoulli's hypothesis is far harder to sustain. Individual prioritarianism is thus "easier" to defend than communal prioritarianism. Yet, the latter remains an attractive alternative that is worth some attention.

[4] This example is a slightly modified version of Rabinowicz's one (2002, 11).
[5] As Rabinowicz notes, this means that Bernoulli's hypothesis applies.

3. The Utilitarian Distribution Principle and the Incoherence of the Priority View

This section states Broome's utilitarian principle of distribution (henceforth, UDP) and explains why, if we accept this principle, we must conclude that the priority view is incoherent. The UDP states that the betterness relation between two prospects is determined by their respective expected goodness. More formally, a prospect A is better than a prospect B if and only if the overall expected goodness g(A) of prospect A is superior to the overall expected goodness g(B) of prospect B and where the overall goodness of any prospect x is

(2) $g(x) = g_1(x) + \ldots + g_n(x) = \sum_{i=1}^{n} g_i(x)$, with i = 1, …, n the members of the relevant population.

The UDP thus states the overall goodness of a prospect is the unweighted sum of individual goodness. Broome establishes the utilitarian distribution principle axiomatically.[6] Five sets of axioms or principles are required to reach equation (2):

1) the individual betterness relation satisfies the axioms of expected utility theory;
2) the general betterness relation satisfies the axioms of expected utility theory;
3) the strong principle of personal good;
4) the rectangular field assumption;
5) Bernoulli's hypothesis.

The first four assumptions are constitutive of the interpersonal addition theorem, which itself is a version of Harsanyi's aggregation theorem (Harsanyi 1955).[7] The interpersonal addition theorem establishes that the general betterness relation between prospects can be *represented* by an expectational utility function defined as the sum of expectational utility functions *representing* the individual betterness relation (Broome 1991, 202). The first two assumptions guarantee three properties regarding the structure of the individual and overall betterness relations: first, they establish that the betterness relations form an *ordering*, *i.e.* a reflexive, transitive and complete ranking of all prospects. Second, they ensure separability across states of nature, *i.e.* the betterness relation between two prospects can be determined by comparing their respective outcomes in each state of nature. Finally, they allow representing the individual and overall goodness of a prospect through an expectational utility function, *i.e.* the probability weighted sum of the utilities attached to each outcome. The principle of personal good applied to prospects has already been discussed above. It guarantees that the general betterness relation supervenes on the individual betterness relation and more generally that

[6] It should be noted that Broome has altered his terminology between *Weighing Goods* (1991) and *Weighing Lives* (2004). In the former, he refers to the combination of the five sets of axioms or principles as the "utilitarian principle of distribution". In the latter, the same combination is referred to as the "interpersonal addition theorem" (2004, 133). In *Weighing Goods*, the interpersonal addition theorem corresponds only to the conjunction of the first four set of axioms. I keep Broome's original terminology, since it helps to single out the role played by Bernoulli's hypothesis in the demonstration leading to the utilitarian conclusion.

[7] Unlike Broome, Harsanyi applied his theorem to the aggregation of preferences rather than goodness. Moreover, in his axiomatic proof, Harsanyi relied on a version of the Pareto principle. As I note in footnote 3, this requires another tacit assumption, *i.e.* that individuals must have a common prior over the probability distribution of the states of nature.

overall goodness is separable across persons. Finally, the rectangular field assumption is a technical assumption. It states that the betterness relations must range across all the possible prospects and outcomes resulting from the combination of the various dimensions (*i.e.* states of nature and people).

Each of these assumptions can be disputed on its own ground. I have already stated that the principle of personal good is seen as controversial by some when applied to prospects. Depending on its application, the rectangular field assumption may imply that the betterness relation must range over meaningless or impossible prospects. The first two assumptions establishing the coherence of the general and individual betterness relations are also open to criticism.[8] However, there are also strong arguments in their favor and to discuss the pros and cons of these assumptions is well beyond the scope of this study. Thus, I assume that we can accept these first four assumptions and thus the interpersonal addition theorem. However, these assumptions are necessary but not sufficient conditions for the UDP to hold. Indeed, the interpersonal addition indicates that the betterness relations can be represented by expectational utility functions. But it does not say that they *must* be represented by such functions. As an illustration, consider the overall goodness of the prospects *x* and *y* of figure 2.

Figure 2

	x			y	
	s_1	s_2		s_1	s_2
i	5	5		2	6
j	7	3		6	6

Assume that assumptions 1 to 4 of the interpersonal addition theorem are satisfied. Then, it follows that we can express the goodness of *x* and *y* by the following utility functions:

$$\text{(3a)} \quad U(x) = g_i(x) + g_j(x) = \frac{1}{2}(5+5) + \frac{1}{2}(7+3) = 10$$

$$\text{(3b)} \quad U(y) = g_i(y) + g_j(y) = \frac{1}{2}(2+6) + \frac{1}{2}(6+6) = 10$$

These functions indicate that prospect *y* and prospect *x* are equally good overall. However, suppose that we use any increasing transform of the function *g* to represent the individual betterness relation. For instance, consider the function $v_i = w(g_i) = g_i^2$. The function *v* represents the same ordering of prospect than *g* according to their individual goodness. If we use the function *v* rather than the function *g* as an input in the social welfare function representing the general betterness relation, then (in order to preserve the ordering over prospects) we must have:

[8] The transitivity axiom is a recurrent target for some philosophers. See for instance Rachels (2005) and Temkin (2012). The independence axiom, which guarantees that states of nature are separable, has also been criticized both as a criterion of individual rationality but also as a criterion for moral judgment (Diamond 1967). I return on this point in the last section.

(4) $$V(x) = \sqrt{w(g_i(x))} + \sqrt{w\left(g_j(x)\right)} = \sqrt{v_i(x)} + \sqrt{v_j(x)}$$

Equation (4) states that overall goodness is a nonlinear combination of individual goodness, *i.e.* the sum of the square roots of individual goodness. The remarkable point is that now we can distinguish the *amount* of good corresponding to a prospect x for a person i as measured by $v = w(g)$ and how this good *counts* in the overall goodness of x (the square root of v). This is precisely the distinction needed by individualistic prioritarianism. Thus, the interpersonal addition theorem is not sufficient to vindicate utilitarianism: it does not show that the ordering of prospects according to their goodness is necessary a utilitarian one. It is the reason why Bernoulli's hypothesis is necessary to give the interpersonal addition theorem a utilitarian meaning. A formal statement of Bernoulli's hypothesis is the following (see Broome 1991, 142):

> *Bernoulli's hypothesis*: A prospect A is at least as good as a prospect B for a person if A's expected goodness is at least as great as B's.

The acceptance of Bernoulli's hypothesis has a crucial implication: it allows one to restrict its attention to the family of utility functions that represent good *cardinally*, *i.e.* the goodness of an alternative is an increasing *linear* transforms of the alternative's utility.[9] Cardinality means that comparisons of utility differences are significant: if the goodness of three prospects A, B and C are represented by cardinal utility functions with the same interval unit, then statements such as "the utility difference between A and B is bigger than between B and C" become meaningful.

More formally, define the goodness $g_i(x)$ for a person i of prospect $x = (o_1, \ldots, o_n)$ where a prospect maps each state of the world into an outcome. If Bernoulli's hypothesis is true, then $g_i(x)$ corresponds to the weighted sum of the goodness of the outcomes:

(5) $$g_i(x) = p_1g_i(o_1) + p_2g_i(o_2) + \ldots + p_ng_i(o_n)$$

Bernoulli's hypothesis implies that for each outcome o_k, we can define a utility function $u_i(o_k)$ such that $g_i(o_k)$ is an increasing linear transform of $u_i(o_k)$. By extension, the same applies to prospect x. Now, consider what the preceding means regarding the general betterness relation. The consequence of Bernoulli's hypothesis is to restrict the family of utility functions that can represent goodness. Denote as $\boldsymbol{U}$ the set of all utility functions that can represent a particular ordering over alternatives according to their goodness. The interpersonal addition theorem states that overall goodness can be represented as the sum of individual goodness. If Bernoulli's hypothesis is true, only the set $\boldsymbol{U^*}$ of cardinal utility functions can represent individual goodness. Indeed, the property of cardinal utility functions is to make utility differences meaningful and this property is needed to represent expected goodness. Then, any

[9] For any utility function u and a goodness function g, g is an increasing linear transform of u if $g = a.u$ with $a > 0$. Actually, Bernoulli's hypothesis is less restricting since it allows one to consider also all positive *affine* transforms of utility functions, *i.e.* $g = a.u + b$ with $a > 0$. Indeed, because the ratio of utility differences is also preserved under any positive affine transformation, it does not matter if all persons' goodness is not measured form the same zero point. Without any loss of generality, I make in the rest of the paper the simplifying assumption that every person's goodness has the same zero point.

individual goodness function $g_i(x)$ is necessarily itself a member of $\boldsymbol{U^*}$ and is thus a positive linear transform of an individual utility function $u_i(x)$. Any function in $\boldsymbol{U^*}$ will preserve the ordering represented by $g_i(x)$ but, since all members of $\boldsymbol{U^*}$ are increasing *linear* transforms of each other, they will also preserve the ratios of utility differences between alternatives. If individual good is only represented by functions pertaining to $\boldsymbol{U^*}$, it follows that, according to the interpersonal addition theorem, the general betterness relation between prospects is necessarily determined by a sum of expectational utility functions measuring individual goodness cardinally. As a consequence, the overall goodness of a prospect corresponds to the sum of any common linear transform of the goodness of the prospect for each person and is thus linear in individual goodness.[10] Actually, this is the UDP.

It may now be easier to understand why prioritarianism such as expressed in equation (4) is incoherent if we accept Bernoulli's hypothesis. For Bernoulli's hypothesis to be true, utility differences must be meaningful. In particular, utility differences *ratios* must be constant through all functional transformations. This restricts the family of relevant utility functions to cardinal utility functions. The combination of Bernoulli's hypothesis with the interpersonal addition theorem thus excludes the use of nonlinear social welfare functions such as (4). Indeed, under nonlinear transforms, utility differences ratios between *individuals* would not be constant. Therefore, complemented with Bernoulli's hypothesis, the interpersonal addition theorem takes a utilitarian meaning because (intrapersonal and interpersonal) comparisons of goodness or utility differences now have a sense.

4. Should We Accept Bernoulli's Hypothesis (I): Good as an Independent Quantity

Whether or not Bernoulli's hypothesis is true has thus tremendous ethical implications. In particular, it vindicates utilitarianism against prioritarianism and various forms of egalitarianism. Until now however, I do not have presented any argument indicating that we should accept this hypothesis. In this section, I briefly discuss and dismiss an intuitive argument that reveals to be unsustainable: good may be an independent quantity such that Bernoulli's hypothesis is true.[11]

In a nutshell, the claim is that we can objectively define what good is and that the way good must be measured is one of its intrinsic properties. In other words, we have, it is argued, a prior quantitative conception of good such that Bernoulli's hypothesis applies to the measure of good. We could also say that goodness is by its very nature a cardinal concept such that the

[10] It is worth adding that the UDP also requires that intrapersonal and interpersonal comparisons of utility (or goodness) are possible. If Bernoulli's hypothesis is true, intrapersonal comparisons of utility are by definition possible: once we have chosen to restrict the representation of a person's betterness ordering to cardinal utility functions, the utility differences between alternatives will be preserved through any increasing linear transform. However, this is not sufficient for *interpersonal* comparisons of utility: since the individual betterness relation of each person can be represented by a family of cardinal utility functions with different unit interval, utility differences between persons are not directly comparable. Individual utility functions must be *co-cardinal*: their unit interval must be the same and the same increasing linear transform must be applied to all utility functions.

[11] Weymark (2005, 27-30) seems to attribute this argument to Broome (1991, 146-147). However, I think that Weymark wrongly interprets here Broome's defense of Bernoulli's hypothesis. See the next section where I present the Broome-Harsanyi argument for Bernoulli's hypothesis.

goodness of a prospect necessarily corresponds to the weighted sum of its outcomes' goodness, or any common linear (or affine) transformation of them. Despite its intuitive appeal, this argument has obvious weaknesses. Firstly, the argument that we have a prior quantitative conception of goodness is meaningless unless we are able to define precisely what goodness is. In a Rawlsian perspective, we could for instance associate goodness to primary goods. But, and this is a second problem, this hardly settles the measurement issue. There are many ways we could measure the quantity of good, even if goodness is well-defined. Measuring goodness requires the definition of a relational structure indicating how two or more "objects" should be combine and then fixing a functional relation mapping this combination into some real number. However, this can be done in many ways such that the same set of objects can be measured in different manners.[12]

Finally, assume that we can bypass the above difficulties. We have a well-defined notion of good as well as a unique metric to measure it. Bernoulli's hypothesis still does not follow. Assume that a person is presented with the two following prospects (see figure 3):

Figure 3

A		B	
s_1	s_2	s_1	s_2
5	5	1	11

Suppose that both states of nature are equiprobable. According to Bernoulli's hypothesis, prospect B is better than prospect A. But if good is an objective quantity that can be independently measured, it is hard to tell why one should accept Bernoulli's hypothesis in this case. Indeed, prospect B is better than prospect A only for risk-neutral agents. But risk-neutrality toward goodness seems not to be more a rational requirement than risk-neutrality toward money is. It seems perfectly reasonable to prefer the sure prospect A that guarantees a moderate but significant quantity of good to prospect B where there are chances to end up with almost nothing. More generally, a quantitative conception of goodness has the odd implication that utilitarianism is true only if people are rationally required to be risk-neutral toward good. But it is generally agreed that risk-attitudes are outside the scope of rationality since whether one is risk averse or risk neutral depends on his preferences regarding risky prospects. Since decision theory (including expected utility theory) does not set constraint on the *content* of preferences, it cannot define what a rational attitude toward risk is. Another possibility would be to argue that individuals are *morally* required to be risk-neutral toward good. This seems implausible because one-person decision problems as the one of figure 3 are probably outside the scope of morality. But even if they were, it remains to make a convincing argument that one *ought* not to prefer the sure prospect in figure 3.

[12] See Weymark's (2005) discussion of measurement theory.

5. Should We Accept Bernoulli's Hypothesis (II): The Broome-Harsanyi Argument

Bernoulli's hypothesis cannot be backed up by the claim that good is a well-defined and independently measured quantity. However, there is another way to defend it by arguing that good as a quantitative notion is defined in exactly the same way utility is in expected utility theory. It follows that once we have measured the *amount* of goodness associated to an alternative, we have also determined how it *counts* overall. I call this defense the Broome-Harsanyi argument since while Broome makes it explicit (Broome 1991, 142-8, 216-218), Harsanyi argued along the same lines in his article "Nonlinear Social Welfare Functions" (Harsanyi 1975).[13]

The Broome-Harsanyi argument starts from a recognition of the objections made in the preceding section regarding goodness: firstly, it is quite unlikely that we have a well-defined quantitative notion of goodness and it is even more unlikely that this notion is arithmetical; secondly, even if good is an arithmetical quantity, it can be quite rational to prefer an alternative whose expected goodness is lower than the expected goodness of another because of risk aversion (Broome 1991, 144). This is a significant objection: if individuals' utility functions are nonlinear in personal goodness, then social (collective) utility will also be nonlinear in personal goodness. Broome and Harsanyi argue otherwise: our quantitative notion of goodness is *derived* from the operation of weighing across states of nature [(Broome 1991, 147); (Harsanyi 1975, 320)]. Let me explain how. If a person's preference ordering or betterness relation satisfies the axioms of expected utility theory (in particular, completeness, transitivity and strong independence), then it is possible to assign a real number to any outcome in the following way (see Binmore 2009). First, we arbitrarily assign a number to two outcomes. For the sake of simplicity, assume that we can identify the outcome the person prefers the less (or which is the worse for her) L and the outcomes she prefers the most (or which is the better for her) H. We decide to assign the following utility numbers to each outcome: $u(L) = 0$ and $u(H) = 10$. Now, for any outcome M which is better than L and worse than H, we can assign a utility number by asking for which probability p the prospect $P = [H, p; L, 1\text{-}p]$ is equally good than M. For instance, say that M is equally good for the person than P when $p = 0{,}7$. Then, we have $u(M) = 7$. The quantitative measure of the utility of M is derived by the weighing across the states of nature. Indeed, M is a sure prospect such that, if it is chosen rather than P, the loss of utility from H to M that occurs with probability p (here, 0,7) is exactly counterbalanced by the gain of utility from L to M which occurs with probability 1-p.

Now, we can go further. Imagine that we have four outcomes H, L, M and N and suppose that the following utility numbers have been assigned: $u(H) = 10$, $u(L) = 0$, $u(M) = 7$ and $u(N) = 2$. Compare now the following two prospects where $p_1 = p_2 = ½$:

[13] As I have already indicated, Harsanyi was not concerned with goodness since he defined utility as the degree of preference satisfaction. Harsanyi's argument directly applied to utility. But actually, if Bernoulli's hypothesis is true, good and utility is the same thing.

Figure 4

	s_1	s_2
A	H	L
B	M	N

Clearly, prospect A's expected utility is higher than prospect B's. Could we argue however, that one should prefer B over A because the former has a smaller variance than the latter?[14] This claim, known as the "utility-dispersion argument" (Harsanyi 1975, 320), is actually totally mistaken: "This is a completely wrong interpretation of the utility notion… The principal result of utility theory for risk is that a linear utility index can be defined which reflects completely a person's preferences among risky alternatives" (Luce and Raiffa 1957, 32). This quote of Luce and Raiffa makes it clear why risk-aversion cannot be invoked as it was the case for goodness in the preceding section: the utility index which we have defined through the procedure described in the paragraph above already incorporates the person's attitude toward risk. As soon as the preference or betterness relation satisfies the axioms of expected utility, preferences or betterness over prospects can be represented by expectational utility functions defined as the probability-weighted sum of the outcomes' utilities.

We can go further. According to the Broome-Harsanyi argument, we can express the fact that A is better than B differently: it is equivalent to say that the utility difference between H and M is greater than the utility difference between N and L, *i.e.* $u(H) – u(M) > u(N) – u(L)$. This can be deduced without looking at the actual utility numbers. Define a third prospect C = [M, ½; L, ½] and ask what is the best thing to do: substitute H for M, or substitute N for L. Obviously, it is better to withdraw M for H if the utility difference between H and M is greater than between N and L. But the former leads to prospect A, the latter to prospect B. Therefore, A is better than B only if the change from M to H is better than the change from L to N. The Broome-Harsanyi argument thus claims that weighing across states of nature makes intrapersonal comparisons of utility possible and meaningful. Indeed, any increasing linear (or affine) transform of the utility function u will preserve the ratio of utility differences. Assume for instance that we define a function $v = 2u + 1$. It is easy to see that $(u(H) – u(M))/ (u(N) – u(L)) = (v(H) – v(M))/ (v(N) – v(L)) = 3/2$. As a consequence, it seems that we have derived a metric measuring utility cardinally.

A question remains though. Is this the case that among all the increasing linear transforms of u and v figures the goodness function g representing the betterness relation among prospects? Actually, the answer is definitely "yes" since nothing prevents us to call the function that has been constructed above a "goodness function" g. It is perfectly sensible to argue that utility measures goodness, or more exactly that we measure goodness in exactly the same way we measure utility, *i.e.* by weighing across states of nature. Indeed, this is the core of the Broome-Harsanyi argument for Bernoulli's hypothesis. Then, the expected utility theorem also applies to goodness. But this claim can be strengthened further by noting that we can

[14] Put in terms of betterness, could we argue that B is better than A because the former has a smaller variance in terms of goodness than A?

derive the same metric by weighing across *persons* rather than states of nature (Broome 1991, 215-217). Consider the following two prospects E and F:

Figure 5

	Persons	
	i	*j*
E	100$	150$
F	50$	250$

Here, a prospect maps a *person* (rather than a state of nature) into an outcome, where without loss of generality an outcome is here defined as a prize in dollars. According to the interpersonal addition theorem, to determine whose prospect is the best, we must compare their expected utility, *i.e.* [u_i(100$) + u_j(150$)] and [u_i(50$) + u_j(250$)].[15] Suppose, as an illustration, that u_i(100$) = 2, u_i(50$) = 0, u_j(150$) = 7 and u_j(250$) = 10. Then, prospect E is the same as prospect B and prospect F is the same as prospect A above. It follows then that F is better than E. This means that as we compare prospect E to F, we give more weight to *j*'s gain of 100$ (from 150$ to 250$) than to *i*'s loss of 50$. This is reflected in the utility numbers I have assigned. Alternatively, assume that E and F are equally good, *i.e.* [u_i(100$) + u_j(150$)] = [u_i(50$) + u_j(250$)]. It follows then that [u_i(100$) - u_i(50$)] = [u_j(250$) - u_j(150$)]. In this case, we give equal weight to *i*'s 50$ loss than to *j*'s 100$ gain. But then, that also implies that the utility numbers ascribe to each money prize and to each person are no longer the same as before. Weighing across persons and weighing across states of nature are thus two different ways to ascribe utility values to outcomes and prospects. The interpersonal addition theorem indicates that, in both cases, the overall goodness of a prospect (its expected goodness) is the sum of individual utilities which have been determined *either* by weighing across persons or by weighing across states of nature. Since utility is a cardinal measure of goodness, it follows that overall goodness can be determined both ways.

One may wonder whether the way we weight outcomes is neutral or not, *i.e.* does both way of weighing lead to the same conclusion regarding overall goodness. The interpersonal addition theorem brings a positive answer. Consider the following example:

[15] Note that the subscripts are required here since we cannot assume that a same quantity of money brings the same amount of utility or good to two individuals. Moreover, we have to assume the possibility of interpersonal comparisons of utility for the addition to make sense.

Figure 6

Prospect G				Prospect H		
Persons/States	s_1	s_2		Persons/States	s_1	s_2
i	100$	200$		*i*	400$	50$
j	100$	200$		*j*	50$	300$

Once again, we assume that both states of nature are equiprobable. Suppose that according to the general betterness relation, G and H are equally good. We can reach this conclusion in two different manners. If we compare both prospects across states of nature, we have

$$(6)\ \big(u_i(100\$) +\ u_j(100\$)\big) - \big(u_i(400\$) +\ u_j(50\$)\big) = \big(u_i(50\$) +\ u_j(300\$)\big) - \big(u_i(200\$) +\ u_j(200\$)\big)$$

If we compare both prospects across persons, we have

$$(7)\ \big(u_i(100\$) +\ u_i(200\$)\big) - \big(u_i(400\$) +\ u_i(50\$)\big) = \big(u_j(50\$) +\ u_j(300\$)\big) - \big(u_j(200\$) +\ u_j(100\$)\big)$$

But, of course, equations (6) and (7) are the same and if G and H are equally good (they have the same expected utility), that necessarily means that comparing utility differences across states of nature or across persons lead to the same result. Both modes of weighing are equivalent.[16] According to Broome (1991, 217), this vindicates Bernoulli's hypothesis: "The very same functions give the weights in weighing across two different dimensions. This very much strengthens the claim that they determine the meaning of quantities of good. The same quantities are what *count* in two different dimensions, and this strongly suggests that they actually *are* quantities of good" (emphasis in original).

6. Is Goodness Cardinal or Ordinal?

The Broome-Harsanyi argument seems to undermine individualistic prioritarianism. But it is not definitive though. Actually, even if it may give one *reasons* to believe that Bernoulli's hypothesis is true regarding good and thus to accept the UDP, this argument is not mathematically supported. This has been established by Weymark (1991) in the context of a discussion of Harsanyi's utilitarian theorems and the same author has reiterated the same point against Broome's defense of Bernoulli's hypothesis (Weymark 2005). This should not be intended to mean that the Broome-Harsanyi's argument is wrong, but only that it is less strong than it seems.

The objection against the Broome-Harsanyi argument lies in the following logical remark: even though the expected utility theorem demonstrates that a preference relation over lotteries or prospects can be represented by an expectational utility function if a set of axioms are

[16] I think it is worth insisting that this result holds only if interpersonal comparisons of utilities are possible. Otherwise, it would be meaningless to say that prospect G is equally good (or better, or worse) than prospect H. It is not sufficient that utility functions are cardinal: *i*'s and *j*'s utility functions must have the same unit interval, *i.e.* they must be *co-cardinal*.

satisfied, it does not show that other utility representations are proscribed, even if they are nonlinear in probabilities (Weymark 1991, 264). Denote R any preference or betterness relation that satisfies the axioms of expected utility, *i.e.* R is reflexive, complete, transitive, continuous and strongly independent in states of nature. Assume that U is an expectational utility function representing R over a set of prospects P. Then, $U(x) \geq U(y)$ if and only if xRy for any prospect $x, y \in P$. Since U is expectational, $U(x) = p_1x_1 + p_2x_2 +\ldots+ p_nx_n$ where $x_{1, \ldots,}$ x_n are the various outcomes of the prospect and $p_1, \ldots , p_n$ their associated probabilities. Any increasing linear transform U' of U equally represents R and, as I have noted, preserves the ratios of utility differences. But is U really a cardinal representation of R? Clearly, this has to be the case for Bernoulli's hypothesis to be true, since we already know that the goodness function g is necessarily a linear increasing transform of U. Unfortunately for Bernoulli's hypothesis, there is a family of increasing *nonlinear* transforms V of U that also represent R. For instance, a function $V(x) = \log U = \log(p_1x_1 + p_2x_2 +\ldots+ p_nx_n)$ also represents R, *i.e.* $V(x) \geq V(y)$ if and only if xRy. However, V is nonlinear in probabilities and this has significant consequences regarding the relevance of Bernoulli's hypothesis.

To understand this point, it is important to remember that the utilitarian interpretation of the interpersonal addition theorem depends on the significance of utility differences, both for intrapersonal and interpersonal comparisons of utility. But, unless we restrict the class of possible utility functions to the expectational ones, utility differences will generally not be preserved across all utility representations of R. As an illustration, consider once again the following example (figure 7):

Figure 7

	s_1	s_2
G	100$	200$
H	50$	300$

Figure 6 describes the decision problem faced by person j in figure 5. Suppose that we have been able to assign the following utility numbers: u(300$) = 10, u(200$) = 9, u(100$) = 5 and u(50$) = 2 . Then, prospect G is better than H. Once again, in terms of utility, that means that u(100$) – u(50$) > u(300$) – u(200$). It is easy to check that any positive affine transform of u preserves this inequality. Moreover, the ratio of utility differences (u(100$) – u(50$))/(u(300$) – u(200$)) = ½ will of course remain the same across all positive affine transforms of u. But even though (as I have tacitly assumed) this person's betterness relation satisfies the axioms of expected utility, it is perfectly right to represent it by any other increasing transform of u and, at least for some of them, neither the inequality nor the ratio of utility differences will be preserved (Weymark 2005).

Obviously, this is in complete contradiction with Bernoulli's hypothesis. The latter implies that intrapersonal comparisons of utility are possible and meaningful. I have explained why in the preceding section. From the above inequality, Bernoulli's hypothesis claims that we can construct a prospect where a fair coin toss determines whether one gains 200$ or 50$ and then ask which is best between substituting the 200$ prize for a 300$ prize or the 50$ prize for a 100$ prize. This would give us a metric for measuring the *intensity* of the betterness relation. Once again, this is necessary if we want to give the interpersonal addition theorem a utilitarian flavor. But actually, expected utility theory and the expected utility theorem do not allow such a move.[17] More exactly, additional assumptions are required to justify that we make such intrapersonal comparisons. To understand this point, it is useful to recall that utility indexes are constructed from a betterness or a preference relation between *pairs* of alternatives. As I have explained, we set the utility value of a sure prospect by comparing it with a risky prospect whose components are two arbitrarily chosen outcomes. But Bernoulli's hypothesis requires something much stronger: utility values must be determined by comparing *pairs of pairs of alternatives*.

Actually, there are two ways of sustaining Bernoulli's hypothesis and of arguing that utility functions offer a cardinal representation of goodness: either we arbitrarily consider that only expectational utility functions represent the betterness relation, or we assume that the relevant way of measuring utility differences is by comparing pairs of pairs of alternatives, as above.[18] Both options are questionable and seem to lack a clear justification. This has obvious implications for the relevance of prioritarianism.

7. Individualistic Prioritarianism *versus* Complete Utilitarianism

In any case, Bernoulli's hypothesis is thus not necessarily implied by expected utility theory, despite the fact that Harsanyi (1975) seemed to have argued otherwise.[19] It should be clear that this conclusion undermines at least partly the charge of incoherence against prioritarianism. The claim that the distinction between *amount* of good and how a given amount *counts* is empty and thus makes prioritarianism meaningless entirely relies on the truthiness of Bernoulli's hypothesis. If the latter is wrong, then the interpersonal addition theorem is not utilitarian. Indeed, absolutely nothing proscribes to define a social welfare function that is nonlinear in individual's utility or goodness. The choice of a particular social welfare function then depends on a series of reasons that have nothing to do with the mathematical structure of expected utility theory.

Given Weymark's objection against the cardinality of goodness, do we have still some reason to take Bernoulli's hypothesis as a reasonable assumption? A first one is suggested by Broome (2008) and Risse (2002) and argues for the natural salience of the expectational

[17] More than fifty years ago, Luce & Raiffa (1957, 32) considered this claim to be one of the most common fallacies regarding expected utility theory.

[18] Weymark (2005) attributes the former strategy to Risse (2002) and to Broome (2008) and the latter to Broome (1991). It seems that Harsanyi has also entertained the latter strategy. Weymark (2005, 32) suggests that in this latter case, Broome starts with "an independent quantitative measure of the degree of well-being". As I argue in footnote 11, this is a curious statement given the fact that Broome makes explicit that we do not have a well-defined quantitative notion of goodness.

[19] Broome (1991) is far more prudent and recognizes that the case for Bernoulli's hypothesis is not totally convincing.

representation of goodness and preference. Consider this long quote from Broome (2008, 230-2, all emphasis are mine except the first):

> "Suppose the person prefers history *A* to *B* and history *B* to *C*. But suppose she is indifferent between *B* for sure and a gamble giving her either *A* or *C* at odds of one to two (that is to say, a gamble giving a 1/3 probability to *A* and a 2/3 probability to *C*). In effect, she is willing to accept one chance of making a gain from *B* to *A* in exchange for two chances of making a loss from *B* to *C*. Since she is willing to accept this gamble, the suggestion is that we should take her degree of preference for *A* over *B* to be twice her degree of preference for *B* over *C*. (…) This certainly supplies a workable concept of degree of preference. I shall call it the *expectational* concept. There are alternatives. Any increasing transform – the square, for instance – of utilities measured this way provides a rival concept of degree. But there is something to be said for the expectational concept as opposed to these others. (…) The expectational concept of degree is the most *natural*, but it is not forced on us by preferences alone. Preferences by themselves do not determine a concept of degree. The expectational concept is derived from preferences together with an idea of *naturalness*. We have a reason to prefer the expectational concept of degree of preference to others: It is more *natural*. This reason carries over to expectational degrees of goodness".

The expectational conception of degree of good or preference is argued to be more *natural* than any other alternative conception. At face, this is arguably plausible. The expectational representation of good has some kind of salience or prominence because of the apparent analogy with the measurement of weight or heat. Weight and heat are indeed measured cardinally and the use of an analogous metric for good or preference easily comes to mind. This can be seen as some kind of convention. Others conventions about measurement of good are available but they are less prominent, either because they do not have obvious analogue in other domains of measurement or because they are less convenient to use. Is natural salience a sufficient reason to choose a particular metric lending support to normative and ethical claims? Actually, there are several strong arguments to consider salience as a valuable and significant reason to accept assumptions and conclusions in the scientific domain (Sugden 2011). It is also highly probable that some (or most) of our ethical beliefs and norms have evolved because they possess some form of salience. In other words, I think that we should recognize that the salience of the expectational representation provides us with a reason to accept Bernoulli's hypothesis.

However, this reason cannot be definitive. Other reasons counterbalance it. The fact that other metrics are equally mathematically correct is one of them. Another one is that it seems perfectly reasonable to consider that the Priority View is also salient in its own way. Indeed, many philosophers and economists reject utilitarianism not for the technical reasons that I am considering here, but more simply because it goes against their intuitive ethical judgments. These judgments do not come from nowhere and they probably have naturalistic and evolutionary origins (Binmore 1998) that make them also salient. All in all, it seems clear that the salience argument is not sufficient to fully vindicate Bernoulli's hypothesis. Actually, the problem is that it is hard to see why an ethical theory should depend on the conventional choice of a particular metric of goodness, even if it is the most natural. Many philosophers would argue otherwise: the metric should be chosen such as to support the ethical theory that one finds the most salient. I fail to see why one claim should have priority over the other.

I have already hinted at the second argument in section 5: the fact that weighing across states of nature and weighing across people lead to the same function and thus indicate that a given amount of good count the same in both dimensions (Broome 1991, 217). But it was also noted in the preceding section that actually this conclusion is reached through comparisons between *pairs of pairs* of alternatives that are not licensed as such by expected utility theory. I do not intend to mean that measuring goodness in this way is necessarily wrong. But it requires some justification beyond the mathematical apparatus of decision theory.[20] The justification in question is supposed to be that the interpersonal addition theorem shows that weighing across two different dimensions (states of nature and people) converge toward the same quantity. Clearly, for this justification to have some force, one must give some credence to the interpersonal addition theorem and its axioms. Since in this article I have assumed from the beginning that the theorem was valid and relevant, to reject it now is not an option. The point is then the following: if we compare any two prospects A and B along a well-defined betterness relation satisfying the expected utility properties and the principle of personal good, then the *degree* according to which A is better than B is the same whether prospects are compared across states of nature or across people. We thus have a cardinal metric to compare prospects. Of course, this is no longer true if we use some nonlinear transform of the utility functions cardinally measuring A's and B's goodness. But, as I understand Broome, this is a strong reason in support of Bernoulli's hypothesis (added to the salience of the expectational representation).

Once again, this argument is not definitive: it strengthens the case for Bernoulli's hypothesis but there remain too many counterarguments to see it as decisive. Firstly, it is relevant only if we accept Broome's ontology regarding goodness, *i.e.* that our quantitative notion of goodness emerges through the operation of weighing across dimensions, analogically to the way economists' define utility. But this ontology is purely formal: it purports to claim what the structure of good is, without giving any hint about the *content* of good. At this stage, Broome's and Harsanyi's respective utilitarianisms separate. A case can be made that informed and rational preferences must be coherent and respect the independence axiom. But matters are far less clear for goodness, at least in what Larry Temkin (2012) calls a "reason-implying sense": if the relevant criterion for comparing two alternatives are a function of the pair of alternatives under consideration, nothing guarantees that we can define a coherent overall ranking of alternatives according to a betterness relation. In this case, our quantitative notion of good cannot be defined in the same way we define utility to represent preferences. I will not dwell further on this issue[21] but this point should be kept in mind when one is weighing the pros and cons for Bernoulli's hypothesis.

Secondly, if we accept that good is measured along the way suggested by Broome, then notions of partiality and impartiality seem to have no place (Hausman 1993). Indeed, it is

[20] See Luce & Raiffa (1957, 32).

[21] In part because, as it should be clear, it threatens the axiom of transitivity of expected utility theory and thus could lead us to discuss the interpersonal addition theorem, which I do not want to do.

meaningless to say that the utilitarian principle treats people impartially because, by definition, one unit of good necessarily *counts* as one unit of good notwithstanding to whom this unit pertains. Contrary to what may appear, this has nothing to do with impartiality because it merges personal benefits with general benefits. Quite the contrary, impartiality requires that we should not give the benefits of some individuals more weight than to the benefits of others. But if Bernoulli's hypothesis is true, the latter sentence is meaningless: the good of individuals who benefit more count more than those who benefit less. This is disturbing and counterintuitive and it may lead us to doubt that this is the way we actually understand goodness (see Hausman 1993, 802).[22] Finally, it could be possible to shift the balance of reasons in favor of Bernoulli's hypothesis if it could be demonstrated that the same quantitative notion of good emerges by weighing through another dimension, namely *time*. Broome entertains this possibility [(1991, 225-240); (2004, 215-223)] but recognizes that it seems highly implausible. Actually, this would require that we form an assumption of separability of time, *i.e.* two alternatives can be compared across their temporally located goodness. If that was the case, then that would imply that we can measure goodness by comparing prospects according to how their outcomes are distributed across times, exactly as we do for persons and states of nature. Then, an "intertemporal" addition theorem would imply that the same quantitative notion of goodness could be derived in three different ways. As a result, we would obtain an extreme form of utilitarianism that Broome (2004, 256-7) calls "complete utilitarianism": the overall goodness of a prospect is equal to the sum of individuals' good which itself is equal to the sum of each person's temporally located good.

Complete utilitarianism requires a corresponding metaphysics regarding personhood to be convincingly defended. Derek Parfit's (Parfit 1984, Part 3) reductionist metaphysics of personhood is probably the most solid account in support of this extreme form of utilitarianism. Parfit argues for the analogy between the dimensions of time and people on the ground that the psychological connection between a same person's temporal selves is not axiologically significant. There is nothing more in being the same person through time that the psychological continuity between temporal selves, and this psychological continuity is not sufficient to justify treating differently intrapersonal and interpersonal distributive relations. Parfit claims that this ontology has two significant implications: firstly, it extends the *scope* of morality. For instance, neglecting his own future is no longer imprudent or irrational but rather immoral because it harms the future selves' wellbeing. Secondly, it tends to downplay the importance of distributive principles: since we generally do not consider that equality in the distribution of goodness across time in the *same* life is morally significant, it follows that equality should not be given more significance regarding the distribution of goodness across persons.

[22] Broome (2004, 91-6) seems to have change his mind on this point. For instance, he writes: "A persons wellbeing or personal good depends only on how things are, seen from a perspective that is personal to her in some sense or other, whereas how her wellbeing counts in general good may depend also on how things are, seen from some sort of external perspective" (93). However, it is not clear how the utilitarian distribution principle fits with "some sort of external perspective". Moreover, even in this latter book, Broome continues to reject the Priority View (Broome 2004, 133).

Broome argues that Parfit's metaphysics is plausible, but whether it supports the separability of time assumption depends on an ethical assumption: that the psychological connections between a person' temporal selves are axiologically non significant. A case needs to be made for this claim and neither Parfit nor Broome offer one. Anyway, it seems highly implausible to appeal to the argument of separability of time to defend Bernoulli's hypothesis and thus utilitarianism: it makes necessary for one to make a commitment on an ethical ground while most prioritarians reject utilitarianism for its implausible ethical implications. Of course, if one is ready to sustain that psychological connections between selves do not matter ethically, then this strengthens the case for Bernoulli's hypothesis. But it also leads one toward a stronger form of utilitarianism. It is not clear that utilitarians can be comfortable with this outcome where they must either endorse complete utilitarianism or grant the possibility of prioritarianism.

8. Prioritarianism, Bernoulli's Hypothesis and the Independence Principle

A last argument regarding the relationship between prioritarianism and Bernoulli's hypothesis has to be considered. This argument is related to one of the drawback that is sometimes attributed to individualistic prioritarianism: the fact that it contradicts the independence principle. This point is developed particularly in McCarthy (2006) against what he calls "*ex ante* prioritarianism", but to my knowledge it has not been articulated in conjunction with the issue of the relevance of Bernoulli's hypothesis.

The strong independence axiom (or independence principle for short) has various formulations but can be stated in the following general way (see McClennen 2009):

> *Independence principle*: Consider any alternatives x, y, z which can be either risky prospects or sure outcomes. For any p such that $0 < p < 1$, we have
>
> if xIy, then $[x, p; z, 1\text{-}p]I[y, p; z, 1\text{-}p]$, where I denotes indifference.

The independence principle implies that any two prospects can be compared across states of nature. In particular, if one is indifferent between the outcomes of two prospects for each state of nature, then the independence principle indicates that one should be indifferent between the two prospects. Equivalently, if a prospect guarantees better outcomes than another one for all states of nature, then the former is better than the latter.[23] The independence principle is needed to grant the expectational form to the utility function representing the betterness relations. Since the independence principle is part of the axioms of expected utility theory, the acceptance of the interpersonal addition theorem means that one also accepts this principle.

Now, consider the following variant of a well-known example discussed by Diamond (1967):

[23] In essence, this is Savage's sure-thing principle.

Figure 8

	x		y	
	s_1	s_2	s_1	s_2
i	3	3	3	1
j	1	1	1	3

The numbers in the matrix indicate the individual goodness g_i and g_j of each outcome. As usual, I assume that states of nature are equiprobable. A quick look at the table is sufficient to understand that prospects x and y should be equally good for someone who accepts the UDP. Applying the independence principle, we can compare the prospects across states of nature. Then, we see that that both prospects give identical outcomes in state s_1. That implies that regarding general goodness, one should be indifferent between both prospects if s_1 obtains. Looking at s_2, we see that even if the outcomes are not identical (the better-off person is not the same one according to the prospect chosen), the other axioms of the UDP imply that neither prospect is better than the other. Therefore, one is led to the conclusion that x and y are equally good.

This conclusion however is not accepted by everyone. Prioritarians in particular have reasons to reject it. Indeed, even though the overall expected goodness of both prospects is the same, their individual expected goodness is not the same for persons i and j: obviously, x is better than y for i, while y is better than x for j. Since there is no general agreement over the individual betterness relation between x and y, individualistic prioritarians who accept the principle of personal good may at first face argue against the utilitarian conclusion by claiming that the general betterness relation should give priority to the (expected) goodness of the least well-off person. Suppose that a prioritarian wants to describe the general betterness relation through the following utility function V,

$$(8)\ V(x) = v_i(x) + v_j(x) = w\big(g_i(x)\big) + w\left(g_j(x)\right) = w\left[\frac{1}{2}\big(g_i(x;s_1)\big) + \frac{1}{2}\big(g_i(x;s_2)\big)\right] + w\left[\frac{1}{2}\big(g_j(x;s_1) + \frac{1}{2}\left(g_j(x;s_2)\right)\big)\right]$$

where $g_i(x,s_1)$ is the goodness of prospect x for person i when state s_1 obtains and with $w(.)$a strictly increasing and concave transform of the goodness function g. Since w is an increasing transform of g, this entails that $v_i(x) > v_i(y)$ if and only if $g_i(x) > g_i(y)$. The same is true for person j. But, because of the concavity of w, it is clear that $V(y) > V(x)$. This leads to the following dilemma: since v is a nonlinear transform of g, either the general betterness relation is nonlinear in individual goodness (as measured by g) or individual goodness (as measured by v) is not expectational. If the former is true, Bernoulli's hypothesis is valid but the independence principle is given up. If the latter is true, then g represents individual goodness *ordinally but not cardinally*.

Suppose for the moment that the numbers in figure 8 represent individual goodness cardinally and therefore that agents compare prospects according to their expected goodness. The prioritarian claim that y is better than x implies a distinction between an amount of goodness (measured by g) and how this amount counts overall (measured by v). The resulting betterness relation exemplifies what McCarthy (2006, 350) calls the "lottery claim principle": everything else equal, it is better to give to people equal chances for some benefit. But, as McCarthy also notes, the lottery claim implies that the betterness relations no longer satisfy the independence principle. This claim is obviously true as illustrated by the current example: prioritarians would like to claim that prospect y is better than prospect x, but I have just concluded above that according to the independence principle both prospects are equally good.

However, there is another alternative: prioritarians may argue that overall goodness is linear in individual goodness *but that individual goodness is not expectational*. In the example above, it happens if we consider that a function $v_i(x) = \frac{1}{2}\left[w_i\left(g_i(x; s_1)\right) + w_i\left(g_i(x; s_2)\right)\right] = w_i(g_i(x))$ is an equally proper representation of individual goodness than g, where $w_i(.)$ is still strictly increasing and concave. In this case, the numbers in the matrix are not meaningful because they cannot be derived through the Broome-Harsanyi argument (more exactly, the numbers may be derived but they have no cardinal meaning).[24] Actually, it means that very few constraints hold in the design of the prioritarian goodness function V: the only constraint is each person's ordinal ranking of prospects through the principle of personal good; beyond this, the prioritarian is absolutely free to represent each person's goodness by any number of its choice. Then, though each person's good is represented by the function g, this function does not have any cardinal meaning and the prioritarian can use any increasing transform v of g to determine overall goodness. The function $V = \sum_i v_i$ is still additively separable though, meaning that the independence principle is not violated.

Therefore, it seems that we obtain some sort of incompatibility result: in general, a prioritarian overall goodness function entails to reject Bernoulli's hypothesis *or* to give up the independence principle. This incompatibility result is of course a straightforward consequence of Broome's and Harsanyi's axiomatic treatment of utilitarianism since we know from the UDP that Bernoulli's hypothesis and the independence principle are among the necessary conditions for a utilitarian social welfare function to obtain. The above analysis also gives a supplementary reason to be skeptical regarding the validity of Bernoulli's hypothesis. Indeed, authors such as McCarthy (2006) consider that the fact that individualistic prioritarianism (or "*ex ante*" prioritarianism as it is sometimes called) is incompatible with the independence principle provides a sufficient reason to reject it. But in virtually all these critiques against prioritarianism, Bernoulli's hypothesis is tacitly taken for granted: personal benefits or individual good are assumed without much justification to be cardinally measurable and to have an expectational form. But these critiques lose much of their strength once we recognize that this tacit assumption is very strong and, actually, maybe too strong. It is then possible to reconcile individualistic prioritarianism and the independence principle simply by rejecting Bernoulli's hypothesis. My point is thus the following: *if* we are ready to grant prioritarianism some ethical significance and plausibility and *if* we are prone to consider the independence

[24] The only other possibility is to assume that we have at our disposal some natural quantitative measure of goodness. We have rejected this possibility in section 4.

principle as a relevant constraint in ethical reasoning,[25] then we *almost* have to reject Bernoulli's hypothesis.

9. Conclusion

The case for Bernoulli's hypothesis remains uncertain. However, what can be said with certainty is that expected utility theory does not offer a logical or mathematical defense of utilitarianism. In particular, it is perfectly consistent to be a prioritarian while acknowledging that the axioms of expected utility theory are solid foundations for a normative theory of the good, at least a far as the structure of the good is concerned. Of course, this conclusion undermines Harsanyi's original claim that if someone who endorses Bayesian rationality in the realm of rationality wants to be consistent, then he must be a utilitarian in the realm of ethics. The impossibility to give logical foundations to utilitarianism on the ground of decision theory seems to be a particular instance of a more general rule: even though the goal to ground morality on rationality has been entertained several times by philosophers and economists (*e.g.* (Binmore 1998); (Gauthier 1986)), I believe that such attempts must ultimately reveal unsuccessful. This is not intended to mean that decision theory and game theory have nothing to offer to ethics; more basically, this indicates that ethics probably does not supervene on rationality.

References

Binmore, Kenneth. G. 1998. *Just Playing: Game Theory and the Social Contract*. MIT Press.

———. 2009. *Rational Decisions*. Princeton University Press.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Wiley.

———. 2004. *Weighing Lives*. Oxford University Press.

———. 2008. Can There Be a Preference-Based Utilitarianism?. In *Justice, Political Liberalism, and Utilitarianism. Themes from Harsanyi and Rawls*, M. Fleurbaey, M. Salles, J.A. Weymark (Eds.), Cambridge University Press, 221-238.

Diamond, Peter A. 1967. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment." *Journal of Political Economy* 75.

Gauthier, D. 1986. *Morals by Agreement*: Oxford University Press.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4): 309–321.

———. 1975. "Nonlinear Social Welfare Functions." *Theory and Decision* 6 (3): 311–332.

Hausman, Daniel M. 1993. "The Structure of Good." *Ethics* 103 (4) (July 1): 792–806.

Luce, Robert Duncan, and Howard Raiffa. 1957. *Games and Decisions: Introduction and Critical Survey*. Courier Dover Publications.

[25] As noted above, this view is rejected by Diamond (1967). See also McClennen (2009) for a critical discussion of the independence principle as a criterion of rationality.

McCarthy, David. 2006. “Utilitarianism and Prioritarianism I.” *Economics and Philosophy* 22 (03): 335–363.

McClennen, Edward. 2009. “The Normative Status of the Independence Principle”. In *The Handbook of Rational and Social Choice,* P. Anand, P.K. Pattanaik and C. Suppe (eds.), 2009, 140-155.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

Parfit, Derek. 1991. “Equality or Priority?”, The Lindley Lecture, University of Kansas, reprinted in *The Ideal of Equality*, M. Clayton and A. Williams, Macmillan (eds.), 2000, 81-125.

Rabinowicz, Wlodek. 2002. “Prioritarianism for Prospects.” *Utilitas* 14 (01): 2–21.

Rachels, Stuart. 2005. “Counterexamples to the Transitivity of ‘Better Than’.” In *Recent Work on Intrinsic Value*, edited by Toni Rønnow-Rasmussen and Michael J. Zimmerman, 17:249–263. Berlin/Heidelberg: Springer-Verlag.

Risse, Mathias. 2002. “Harsanyi’s ‘Utilitarian Theorem’ and Utilitarianism.” *Noûs* 36 (4): 550–577.

Sen, A. 1986. Social Choice Theory. In *Handbook of Mathematical Economics, Vol. III,* K.J. Arrow & M.D. Intriligator (Eds.), North-Holland, 1073-1181.

Sugden, Robert. 2011. “Salience, Inductive Reasoning and the Emergence of Conventions.” *Journal of Economic Behavior & Organization* 79 (1-2): 35–47.

Temkin, Larry S. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford University Press.

Weymark, J. 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In *Interpersonal Comparisons of Well-Being*, J. Elster and J. Roemer (Eds.), Cambridge University Press, 255-320.

Weymark, John A. 2005. “Measurement Theory and the Foundations of Utilitarianism.” *Social Choice and Welfare* 25 (2-3): 527–555.