Review of Louis Narens and Brian Skyrms, The Pursuit of Happiness: Philosophical and Psychological Foundations of Utility[*]

Krister Bykvist and Johan E. Gustafsson

Luis Narens and Brian Skyrms's *The Pursuit of Happiness* is an attempt to take seriously utilitarianism's problem with how to measure happiness. The problem is to find a way to make sense of measurements of happiness so that the utilitarian aggregates of happiness will be meaningful. The first part of the book provides a very useful historical overview of the measurement of happiness in utilitarian theory. As far as we know, this is the first overview of this kind. This part covers Bentham, Mill, Jevons, Edgeworth, and the nineteenth-century psychophysics. It then goes on to present von Neumann and Morgenstern's method of generating cardinal utilities from ordinal preferences over lotteries and Harsanyi's aggregation theorem.[1] The second part provides an overview of modern measurement theory. This overview includes a discussion of modern psychophysics p. 234 and shows its relevance to utilitarianism, which is rarely done. Finally, in the third part of the book, the authors aim at attaching a definite meaning to utilitarianism, but their conclusion is that psychology, neurobiology, and modern measurement theory do not take us very far. These approaches fail to do so, because they fail to deliver a satisfactory solution to the problem of interpersonal comparisons of utility.

The authors suggest two ways forward. The first is to reject the possibility of interpersonal comparisons and change the definition of utili-

[*] Published in *Economics and Philosophy* 40 (1): 233–239, 2024.

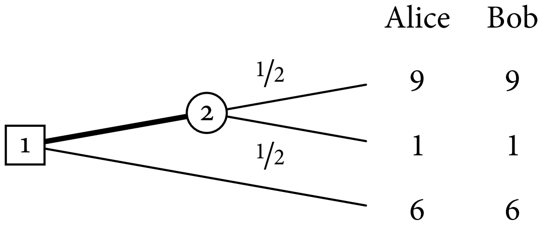
[1] The presentation of von Neumann and Morgenstern's expected-utility theorem has some unfortunate errors, however. On page 66, ordering is defined as just completeness. But Ordering is completeness and transitivity, which is what is required for von Neumann and Morgenstern's theorem. (Likewise, the explanation of Ordering on page 67 makes the same mistake.) On page 67, in the definition of Independence, 'for all a' should be 'for all $a > 0$'. Otherwise, Independence would demand that p is preferred p' only if something is preferred to itself.

tarianism so that it asks us to maximize the *product* of individual utilities rather than the sum, where each individual's utility is measured on a ratio scale.[2] The second is to see interpersonal comparisons of utility as *conventions* rather than matters of fact. The first option is not very attractive, however.

One of many problems (conceded by the authors, pp. 153–154) is that the product approach can only handle cases where everyone has a positive well-being level or everyone has negative well-being, which severely limits not only the applicability but also the explanatory value of the theory. But even if we limit the product approach to fixed population cases where everyone has positive well-being, it runs into problems when we assess risky prospects. We can evaluate such prospects either *ex-post* or *ex ante*. According to *Ex-Post Product Utilitarianism*, we calculate the value of a prospect by first calculating the product of well-being in each final outcome and then taking the expectation of these values. In contrast, according to *Ex-Ante Product Utilitarianism*, we calculate the value of a prospect by first calculating each persons expected well-being and then taking the product of people's expected well-being.

Ex-Post Product Utilitarianism can, as Narens and Skyrms note (p. 158), oppose everyone's expected well-being. Consider

Case One



Here, there the square represents a choice node and the circle represents a chance node. If we go up at node 1, then there is a one-in-two chance that chance goes up at node 2 and Alice and Bob both get a well-being of 9 ; otherwise they both get a well-being of 1. If we go down at node 1, Alice and Bob both get a well-being of 6.

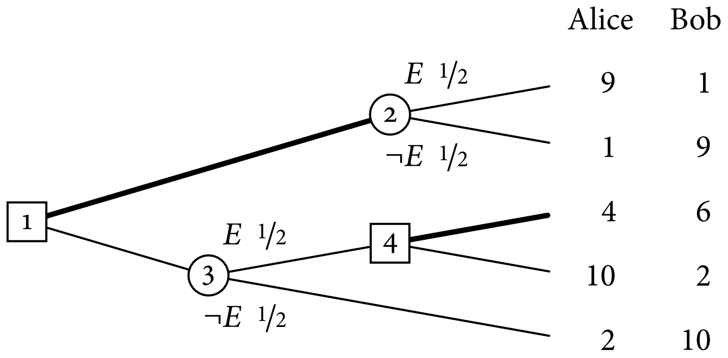
According to *Ex-Post* Product Utilitarianism, the value of going up is equal to $(9\cdot 9)\cdot 1/2+(1\cdot 1)\cdot 1/2 = 41$. And the value of going down is equal to $6\cdot 6 = 36$. Hence, according to *Ex-Post* Product Utilitarianism, it is better

[2] This proposal builds on the earlier paper Skyrms and Narens 2019.

to go up (this recommendation is represented by the thick line). Going up at node 1 gives everyone an expected well-being of $9 \cdot 1/2 + 1 \cdot 1/2 = 5$, whereas going down gives everyone an expected well-being of 6 Hence, at node 1, everyone gets a greater expected well-being if we go up than if p. 235 we go down. Even so, *Ex-Ante* Product Utilitarianism recommends going down.

Ex-Ante Product Utilitarianism avoids this implication in Case One. Nevertheless, it can lead to even worse results in sequential cases. It can be worse to follow its recommendations for everyone, whatever happens, than to follow the opposite recommendations. Consider

Case Two



Here, the there are two chance nodes, which both result in going up if and only if the same one-in-two chance event E happens.

At node 4, going up has a value of $4 \cdot 6 = 24$ and going down has a value of $10 \cdot 2 = 20$. Accordingly, *Ex-Ante* Product Utilitarianism recommends going up at node 4. Taking the prediction that we would go up at node 4 into account with backward induction at node 1, going down gives Alice an expected well-being of $4 \cdot 1/2 + 2 \cdot 1/2 = 3$ and Bob an expected well-being of $6 \cdot 1/2 + 10 \cdot 1/2 = 8$. Hence the value of going down at node 1 is equal to $3 \cdot 8 = 24$. Going up at node 1 gives Alice an expected well-being of $9 \cdot 1/2 + 1 \cdot 1/2 = 5$ and Bob and expected well-being of $1 \cdot 1/2 + 9 \cdot 1/2 = 5$. Hence the value of going up at node 1 is equal to $5 \cdot 5 = 25$. Accordingly, *Ex-Ante* Product Utilitarianism recommends going up at node 1. (The recommendations of *Ex-Ante* Product Utilitarianism are represented by the thick lines.)

But compare this recommendation with doing the opposite of what *Ex-Ante* Product Utilitarianism recommends — namely, to go down at all nodes:

	E happens		*E* does not happen	
	ALICE	BOB	ALICE	BOB
Up at node 1	9	1	1	9
Down at nodes 1 and 4	10	2	2	10

In Case Two — regardless of whether *E* happens — everyone is worse off if we follow the recommendations of *Ex-Ante* Product Utilitarianism than if we follow the opposite of its recommendations. Hence, on both the *ex-post* and the *ex-ante* approach, product utilitarianism can oppose everyone's interest. p. 236

Another limitation of product utilitarianism is that it cannot handle variable population cases. The authors do not see this as a weakness, however. They point out that it blocks Derek Parfit's (1984, p. 388) *Repugnant Conclusion* — that is, the claim that, for every population of lives with very high quality, there is a better and much larger population with lives that are barely worth living. Such claims are meaningless on Narens and Skyrms's approach, since saying that an added life has utility 2 (which would double the total value of the population) is equivalent to saying that it has utility 1/2 (which would half the total value), if we assume intrapersonal ratio-scale comparability (p. 155). But the approach does not escape other counter-intuitive implications in population axiology. It seems intuitive not only that the Repugnant Conclusion is false but also that the reverse is true: that there are some populations of lives with high quality that would be better than a much larger population of lives that are barely worth living, but Narens and Skyrms cannot account for this judgement. Moreover, there seem to be some comparisons between populations with different but the same number of people that only seem to require us to compare miserable lives with happy lives. Consider creating a future with lots of people living extremely happy lives or creating a future with different but equally many lives that are instead extremely miserable.[3] It seem that the mere fact that the future people are happy in the first alternative and miserable in the second should be sufficient to let us conclude that there is more well-being in the former. But Narens and Skyrms's approach does not allow this if we look at the values of the whole worlds (which contains some past or present well-being, positive, negative, or neutral).

The second way forward that Narens and Skyrms propose allows us to make interpersonal comparisons of utility, but understands them as con-

[3] Arrhenius 2009, p. 293.

ventions, or at least as having an element of convention. They then show that how these conventions will evolve for a wide class of dynamic games. The utilitarian rule is wheeled in to help us select among Nash equilibria. These results are interesting and worth delving into. But here we would like to take a step back and ask what it means to say that interpersonal comparisons are conventions.

The authors suggest different versions of this view. One version sees interpersonal comparisons as to some extent involving moral judgements, and a second version sees them as purely conventional, which is the version the authors develop further (pp. 160–161). On this latter view, interpersonal comparisons cannot be true, but we often mistakenly think they can be in those cases where the comparisons help us to select among equilibria in coordination problems. This radical view about the nature of interpersonal comparisons cannot be assessed properly until it is made clear exactly what is meant by 'utility' and 'comparison'. As Broome (1991a) and others have pointed out, the term 'utility' can be used to denote

(a) the underlying empirical feature of a person's life that is supposed to have value for that person, pleasure in the case of hedonism about well-being;

(b) the value for a person of some empirical feature of that person's life; the value a person's pleasure has for the person, if we assume hedonism about well-being; p. 237

(c) the general (impersonal) value of some empirical feature of a person's life; the general value of pleasure in the case of hedonism about general value;

(d) the number (or other mathematical entity) that is supposed to represent any of the above features or values in the measurement of the feature or value.

By 'comparison' we can mean comparison of levels, differences, or ratios of some feature or value. Putting this together, we get $4 \cdot 3 = 12$ different meanings to 'interpersonal comparisons of utility'. Which one(s) do the authors have in mind when they say that these comparisons are conventions? Well, it is not clear. Of course, no one would deny that conventions are involved in assigning numbers to represent any of factors (a) to (c). To use one particular utility function rather than another that captures the same information (comparisons of levels, differences, or ratios) is, of course, purely conventional. Often it looks like the authors are mainly talking about (a), exemplified by pleasure or preference satisfaction. But to say that all comparisons of these empirical factors are purely conventional and all false would be absurd, since no one would deny that we can make true interpersonal comparisons of the following kind:

- If I feel pleasure and you feel displeasure (or feel indifferent), then I feel more pleasure than you do.
- If I love Marmite and you hate it (or are neutral towards it), then I want it more than you do.

These 'fixed points' have to be respected in any measurement of different people's pleasures or attitudes. This is not to say that it is easy to know the fixed points.

The authors' focus, however, is on comparisons of differences of pleasure or preference satisfaction, since this is what classical utilitarianism requires in order to meaningfully talk about maximizing the sum of utilities (in fixed population cases); there is no need to compare attitudinal levels. But, if the fixed points above are accepted and we can make some rough interpersonal comparisons of attitudinal levels, then some interpersonal comparisons of differences follow automatically. For example, if I favour Marmite and am neutral towards margarine and you hate Marmite and favour margarine as much as I favour Marmite, then the difference in my attitudinal levels between Marmite and margarine must be less than the difference in your attitudinal levels between margarine and Marmite. This is an instance of a more general phenomena, which is sometimes called 'ordinal intensity':[4]

	My preference	Your preference
Levels	ordering	ordering
1	x	y
2	y	
3		x

p. 238

In this schematic example, the difference in my attitudinal levels between x and y must be less than the difference in your attitudinal levels between y and x. Or, more simply put, my preference for x over y is weaker that your preference for y over x.

[4] Sen 1976, p. 221.

Since we can establish some limited comparisons of differences given some interpersonal comparisons of levels, it is not correct to say that all interpersonal comparisons of differences are purely conventional and false. So, at most we can say that interpersonal comparisons of pleasures or attitudes are only *in part* conventional and false. Furthermore, since there are some instances where the concept of utility difference can be correctly applied, the concept cannot be incoherent.

If we turn to the notion of well-being, what is *good for* individuals, there are two ways to establish interpersonal comparisons of differences of well-being that are not discussed by the authors. One is to take seriously the idea that interpersonal comparisons are moral judgements about overall general goodness of outcomes but deny this means that they are purely conventional and false. If these judgements can be true, then one could define comparisons of differences of well-being in terms of comparisons of the general value of outcomes. Suppose I am better off in A than in B and you are better off in B than in A, and A is better than B. Then we should say that the difference in my well-being between A and B is greater than the difference in your well-being between B and A. The fact that my well-being counts more than yours for the general value of these outcomes shows that my benefit would be greater than your loss if A were to be chosen over B.[5]

A potential problem with this account is that one might want to distinguish well-being and general goodness more sharply. For instance, it seems meaningful to 'give priority to the worse off' and say that your gain in well-being counts for more than my equally sized gain because you are worse off than me.

A different account allows for this and says instead that when we compare the well-being of different individuals what we are fundamentally comparing are the well-being values of *types* of lives.[6] How well off someone is in a certain type of life does not depend on the identity of the person living the life. If I were to lead your type of life, than the value this life has for you would also be the value this life would have for me. More generally, the value of life L for S is equal to the value of life L for S', for all S and S'. Given this invariance assumption, we can use the von Neumann-Morgenstern approach, now applied to the well-being ranking of lotteries of types of lives, and derive difference comparisons that will

[5] Broome 1991b, p. 220.

[6] Broome 2004, pp. 91–94.

hold both intra- and interpersonally. Suppose L_1 is ranked above L_2 and L_4 is ranked above L_3 in the well-being ordering of types of lives. Then the difference in well-being between lives L_1 and L_2 is greater than difference in well-being between L_3 and L_4, if the well-being value of the lottery $(L_1, 1/2, L_3)$ is greater than that of the lottery $(L_2, 1/2, L_4)$. Armed with this measure of interpersonal comparisons of well-being, utilitarianism and its aggregation method can be given meaning.

Of course, the authors could say that this still relies on conventions, in this case a convention of how to rank types of lives in terms of well-being, and such conventions cannot be true. But this is to take a stand p. 239 on a very controversial meta-ethical issue about the nature of well-being judgements, and they owe us arguments for this radical form of conventionalism. No such argument is presented or hinted at in their book.

We would like to end on a positive note. Even though we have some concerns about the third part of the book, it is important to stress that the others parts are very rewarding. As we pointed out in the beginning, the first and second parts provide a historical survey that neatly summarizes the measurement debate from Bentham to Harsanyi. In addition, the role of psychophysics is explored in the context of utilitarianism.

References

Arrhenius, Gustaf (2009) 'Can the Person Affecting Restriction Solve the Problems in Population Ethics?', in Melinda A. Roberts and David T. Wasserman, eds., *Harming Future: Persons Ethics, Genetics and the Nonidentity Problem*, pp. 289–314, Berlin: Springer.

Broome, John (1991a) "Utility", *Economics and Philosophy* 7 (1): 1–12.

— (1991b) *Weighing Goods: Equality, Uncertainty and Time*, Oxford: Basil Blackwell.

— (2004) *Weighing Lives*, Oxford: Oxford University Press.

Parfit, Derek (1984) *Reasons and Persons*, Oxford: Clarendon Press.

Sen, Amartya (1976) 'Liberty, Unanimity and Rights', *Economica* 43 (171): 217–245.

Skyrms, Brian and Louis Narens (2019) 'Measuring the Hedonimeter', *Philosophical Studies* 176 (12): 3199–3210.