

Please cite as:

Browning, H. & Veit, W. (2021). Does Utilitarianism Need a Rethink? *Preprint*.
Check www.walterveit.com/publications for citation details once published

Does Utilitarianism Need a Rethink?

Book Review of Louis Narens and Brian Skyrms'

The Pursuit of Happiness

Heather Browning & Walter Veit

1 Introduction

Philosophers have typically shown high confidence in their evaluations of Utilitarianism, whether as an endorsement or a disparagement. Rarely, however, has much effort been spent on investigating what utilitarianism *means*. In their recent 2020 book *The Pursuit of Happiness: Philosophical and Psychological Foundations of Utility*, the cognitive scientist Louis Narens and the philosopher of science Brian Skyrms have teamed up to address the question of how utility can be measured and aggregated for the purposes of ethics and policy-making.

The Pursuit of Happiness constitutes a beautiful example of benefits of collaboration between scientists and philosophers. Narens, one of the leading experts on measurement theory in the behavioral and cognitive sciences, and Skyrms, a pioneer in mathematical philosophy with his work on evolutionary game theory and the evolution of conventions, make a perfect team for examining the foundations of utilitarianism. The majority of *The Pursuit of Happiness* is a celebration of the utilitarian tradition, particularly as framed by the work of Jeremy Bentham. Narens and Skyrms provide us with a long and detailed history and theory of utility measurement, across both economics and psychophysics, with their own new ideas appearing only in the last few chapters. Instead of defending a thesis from beginning to end, Narens and Skyrms provide us with an elegant historical tale of the original ambitions Bentham had for the utilitarian project and how this vision came subsequently to be replaced. They finish by proposing a novel approach to utilitarianism that they take to be scientifically feasible and which avoids some of the problems associated with traditional views. Their goal, however, is not to radically revise the utilitarian framework, but rather to take an alternative approach to the problems it was meant to solve all along, i.e.

hedonic comparisons across agents, and their admiration for the utilitarian tradition can be felt throughout their book.

Their book contains three key ideas that are aimed at returning us to Bentham's original goals, and these will be the focus of this review. The first is their proposal for ratio-scaled measurement of utility using preferences over hedonic episodes. The second is their method of aggregation of individual utilities using product utilitarianism. And finally, the third is the setting of individual utility ratios by way of convention. All three are a response to what they take to be the biggest problem in measurement of utility: the problem of interpersonal comparisons. As we will argue, however, the problem is not as insoluble as the authors take it to be, and thus their 'radical' proposed solutions may not be required.

2 Measuring Utility

How can we measure utility? Whereas Bentham and the early utilitarians were hedonists and conceived of utility as a measure of subjective experience, economists at the end of the nineteenth century - such as William Stanley Jevons (1871) in his *Theory of Political Economy* - took direct measurement of hedonic feelings to be impossible and thus argued for revealed preferences as something like a next-best alternative.¹ This was not a denial that hedonic experience is important, but merely an assertion that a science of economics should restrict itself to the directly measurable. In doing so, economics became detached from the original moral motivations of utilitarianism. As Narens and Skyrms put it: “[Jevons] wanted to make economics a science, not ethics. Utilitarian ethics is thrown out the window” (p. 23). In such a picture, it is hardly surprising that preference-satisfaction itself eventually came to be regarded as constituting wellbeing, rather than a mere indicator for good welfare.

In their discussion of how utility can be measured, Narens and Skyrms assess the numerous attempts to find a ratio-scaled measure of utility, and how these approaches have yet to succeed. While Von Neumann and Morgenstern came up with a method of constructing an interval-scaled utility function, using preferences over lotteries, this method would not help us to engage in interpersonal comparisons of utility. For Narens and Skyrms' solution to this problem (product utilitarianism, as will be discussed in the next

¹See also Whalley (2005).

section), they require a ratio-scaled measure. A ratio scale, as well as having cardinal units, also requires a non-arbitrary zero point. Narens and Skyrms propose to achieve this by measuring preferences across episodes.

They describe how for others interested in hedonic utility, such as Bentham and Edgeworth, the meaningful bearers of utility were hedonic ‘episodes’ – measured as intensity \times duration – which can then be summed or integrated. Narens and Skyrms propose that Von-Neumann-Morgenstern representation can be used across episodes, if duration is taken to be analogous to probability (e.g. that an episode with $1/3$ intensity A and $2/3$ intensity B represents a gamble of $1/3$ probability A etc.). This then means that we can use concatenations of durations to create a ‘composite’ episode and so long as intensity is fixed, we get additivity. The ‘null episode’ (zero duration, thus zero value) can fix a zero point for intensity – the intensity such that an episode of some positive duration of this intensity is equal in value to that of no duration. Thus, the unit will be arbitrary but the zero meaningful, and this zero will be the dividing line between pleasure and pain. The arbitrary unit should not be a concern, as Narens and Skyrms argue that they do not correspond to anything in reality, and will be fixed by convention only – a suggestion that will surely strike many in the hedonist tradition as strange. But Narens and Skyrms maintain that more is simply not necessary to make utilitarianism work in practice. This will require some unpacking and leads us to their second idea – aggregation through product utilitarianism.

3 Aggregation

Although Narens and Skyrms argue that we can use the values of episodes to set individual utility, they point out that we still run into a problem of summing hedonic values, because the unit is arbitrary. We can multiply each individual’s utilities by any constant, and this will then give us a different ordering of aggregate results, depending on the constants chosen. This creates the requirement for a different aggregation formula to measure utility and they propose that instead of traditional utilitarianism sums, we instead need to take the utilitarian product. Here we aggregate not by the sum, but by the product of individual utilities – this allows the multipliers (whatever they are) to cancel out on either side of the equation. This sense of an allowed arbitrary conversion of any individual’s utility relative to another’s, comes from their take on the interpersonal comparison problem, as we will discuss

in a later section. If we were able to set the relative utilities in a principled way, this problem would not appear and the case for product utilitarianism would be weaker.

There are several limits to the method of product utilitarianism. The first is that it is limited to aggregation or comparison of options of the same valence, such as distribution questions in which all options give positive utility. For ‘mixed’ episodes containing both positive and negative valence, it is unclear how to proceed. We can aggregate each separately but non-additivity will prevent simple subtraction of pain from pleasure, and the authors conclude: “at this point we see no obvious way to combine group pleasure and pain so that such combinations can be meaningfully compared” (p. 154). This obviously rules out use across a large range of cases with such mixed results. The second problem is that it doesn’t allow for comparison of different groups, as the aggregate products to compare must be from the same individuals – looking at how the aggregate changes with the addition or subtraction of individuals is not meaningful. Again, this will constrain the types of comparisons we can perform, particularly if looking at problems in population ethics that require assessment of the value of different populations. It will also not allow for the inclusion of chance over future episodes, as the results will change depending on whether this multiplier is applied before or after aggregation. They admit this, and propose that one will need to give up some tenet or assumption of choice – either coherence or Pareto optimality – or go back to sum utilitarianism, and all the problems it has that product utilitarianism aims to solve.

These are big bullets to bite and concerns that will surely strike many readers of Narens and Skyrms’ proposal as too easily dismissed. But trade-offs in modeling are inevitable (see Weisberg 2013; Veit 2019), and we must keep in mind that their goal is merely to provide a useful method of aggregating utility for the moral purposes originally envisioned by the early utilitarians. Despite the limitations of their assumptions, Narens and Skyrms maintain that product utilitarianism is superior because it is a feasible option in the face of the problem of interpersonal comparison - the aggregation of the products of individual utilities providing a way of reaching answers without requiring interpersonal comparisons. As we will discuss, how convincing their case is will depend on how pressing one takes this problem to be.

4 Convention

In the penultimate chapter, Narens and Skyrms provide an alternative solution, one that would allow for utilitarian aggregation via sum, and this is through use of convention. That is, they think we can use convention to set relative units of utility across individuals, and thus proceed to make comparisons in this manner. There are several ways we could do so, such as using moral judgements or social conventions. They suggest that this may have been how people were actually doing it all along, without being explicit - that all our judgements regarding relative utilities have been implicit conventions. They instead suggest a process of explicit choice of convention by deliberation - communication between actors to allow for setting of convention regarding choice of units, drawing on their work on co-ordination problems between multiple actors.

They caution that this will be in some sense arbitrary, i.e. there will not necessarily be a unique stable equilibrium and we should not take the outcomes of deliberation to represent an objective choice of unit. Some arbitrariness is unavoidable, but its effects can be overcome if we take this conventional approach. The equilibrium will then depend on the intentions of the agents, such as their level of goodwill towards or exploitation of others. The actors will then converge on an agreement as to the correct multiplier of units from one to the other. The solution will be the same with multiple actors, just more complex. Once relative units are set, we can then rescue sum utilitarianism; all that is required are these agreed standards of comparison. Here, they emphasise that “interpersonal comparisons then have a moral status rather than a factual one” (p. 171) - they represent how we want to consider one another rather than any facts about our different psychologies. Once the convention has been set, we can then empirically determine the number of units of pleasure (or pain) each individual gets from a particular resource, and convert these onto a common scale to perform the necessary aggregations and comparisons. We are thus left with two possible solutions - use of conventions to set multipliers for utilitarian sums, or use of utilitarian products to avoid the need for such comparisons at all.

5 The Problem of Interpersonal Comparison

In the end, the proposals we have discussed here are responses to the problem of interpersonal comparison - that is, how we can meaningfully compare the utilities of different individuals for utilitarian calculations. Which solution one prefers will then depend on how strongly one feels the pull of this problem. Naturally, this is one of the oldest and most widely-discussed problems in utility measurement. It is a large part of the explanation of the move in economics to ordinal utility functions, and Pareto-derived decision procedures. There have been a range of proposed solutions, such as use of imaginative empathy or empathic preferences, where relative utilities are set via A's conception of B's utility (and vice versa) (Harsanyi 1955; Binmore 2009), however Narens and Skyrms are rightly sceptical of such methods as they will run into problems when these judgements disagree. Justifying through use of similarity assumptions - taking people to be roughly the same - will also have issues when people are genuinely different, and particularly if we move outside just the human case to comparisons across species (Browning 2020).

It is clear that Narens and Skyrms do not believe it to be possible to arrive at an empirical solution - indeed, they appear to take it to be the case that no such solution exists. They claim that there is no empirical truth of the matter beyond convention: "People just believe that there is a truth of the matter, which there isn't, and blunder around until they fall into a convention, which they believe is the truth" and that "multiple conventions are possible, and none is truer than another" (p. 161). This is an idea that is inspired by Ken Binmore's work on interpersonal comparisons (Binmore 2009). Unlike Binmore, however, they seem to be taking the problem to be metaphysical rather than epistemic - that there is *no* fact of the matter regarding relative utilities, not just that our current methods of measurement may not allow us to find out what they are.

The strong metaphysical claim is a difficult one to make sense of - that there is really no truth of the matter regarding the relative intensity of different experiences, regardless of our measurement capabilities. It seems counter to our intuitive judgements about our ability to assess the relative hedonic intensity experienced by different individuals, though they would argue that this is just an assumption based on our tendency to want to find such a scale. This is a scepticism we do not share - largely for reasons regarding the biological underpinnings of hedonic experience, and the anatomical

and evolutionary similarities between individuals that could provide natural grounding for facts about relative utility (Browning 2020). If it is possible, at least in principle, to empirically determine relative utilities, then utilitarian sums can be rescued once such methods become available and their more radical solutions may not be necessary.

6 Conclusion

In conclusion, readers will appreciate the rich and illuminating discussion of the utilitarian tradition from Bentham to the present. Narens and Skyrms are correct that philosophers after Bentham have given too little attention to the measurability of utility as crucial for use within ethics and politics. Too often has it been treated as something that could at least in principle be determined without any examination of whether or how this might work in practice, and how we measure will impact our assessments of what ought to be done.

Ultimately, however, we think that their apparent denial of a reality for an objective weighting of individual utility is premature. Their case builds on the premise that we have to have an alternative *if* it's impossible. From this perspective, their proposal can be seen as an attempt to recuse utilitarianism itself. But we are not so sure that Bentham-style hedonist utility is such a lost cause after all. While their book draws on psychophysicists and economics for potential ways to measure hedonic utility, the attempts of neuroeconomists to provide such a solution is given only a relatively cursory glance. This is unfortunate because it is within science, and not within philosophy that we would expect to find a successful naturalization of Bentham's hedonic 'common currency'.

In their 2004 call to bring neuroscience to economics, Camerer et al. cite Jevons's (1871) famous line that he "hesitate[s] to say that men will ever have the means of measuring directly the feelings of the human heart. It is from the quantitative effects of the feelings that we must estimate their comparative amounts", as something that we can now overcome (p. 556).² They criticize his skepticism regarding the investigability of the subjective maintaining that "Jevons was wrong. Feelings and thoughts *can* be measured directly, because of recent breakthroughs in neuroscience" (p. 556). *This* assertion was also premature. But we should not downplay the possibilities,

²See Jevons (1871, p. 11).

and neuroeconomics, animal welfare science, and the study of affect-based decision-making have made great strides to make pleasure and pain measurable quantities. Unfortunately, Narens and Skyrms's discussion of this literature is brief and we fear that many of the scientists engaged in the project of naturalizing wellbeing as a biological phenomenon will be left unimpressed with the cursory glance given to their work. Yet, Narens and Skyrms are certainly right that there is no guarantee this naturalist project will succeed. What *The Pursuit of Happiness* provides us with is an elegant intermediary or at least next-best solution to the problems Bentham originally wanted to solve and remain at the core of many of our ethical and political struggles.

References

- Binmore, K. (2009). Interpersonal comparison of utility. In H. Kincaid and D. Ross (Eds.), *The Oxford Handbook of Philosophy of Economics*, pp. 540–559. New York: Oxford University Press.
- Browning, H. (2020). *If I Could Talk to the Animals: Measuring Subjective Animal Welfare*. Ph. D. thesis, Australian National University. <https://doi.org/10.25911/5f1572fb1b5be>.
- Camerer, C. F., G. Loewenstein, and D. Prelec (2004). Neuroeconomics: Why economics needs brains. *scandinavian Journal of Economics* 106(3), 555–579.
- Harsanyi, J. C. (1955, August). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63(4), 309–321.
- Jevons, W. S. (1871). *Theory of Political Economy*. London: MacMillian & Co.
- Narens, L. and B. Skyrms (2020). *The Pursuit of Happiness: Philosophical and Psychological Foundations of Utility*. Oxford University Press.
- Veit, W. (2019). Model pluralism. *Philosophy of the Social Sciences* 50(2), 91–114. <https://doi.org/10.1177/0048393119894897>.
- Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. New York: Oxford University Press.
- Whalley, J. (2005). Rationality, irrationality and economic cognition. *Technical report. CESifo Working Paper No. 1445*.