Evaluating Time-Continuous Action Alternatives from the Perspective of Negative Utilitarianism

A Layered Approach

Erich H. Rast
Philosophy of Language Institute (IFL)
Universidade Nova de Lisboa, FCSH
Lisbon, Portugal
erast@fcsh.unl.pt

Abstract—**A layered approach to the evaluation of action alternatives with continuous time for decision making under the moral doctrine of Negative Utilitarianism is presented and briefly discussed from a philosophical perspective.**

Key words-negative utilitarianism; decision making; Maximin; Expected Utility Theory; Precautionary Principle;

I. INTRODUCTION

In theories of rational decision making such as Expected Utility Theory (EU) [1, 2, 3] and Subjective Expected Utility Theory [4], as well as in related fields such as Consumer Theory in classical economics, preferences among alternatives are commonly represented by a single-agent utility function that aggregates multiple attributes of each alternative into a real number between 0 and 1. In the simplest model of decision making under risk, EU recommends that alternative whose aggregate value times the probability of its occurrence is maximal. In a time-continuous setting one consequence of this approach is that any period of negative value ('disvalue': pain, disutility, displeasure, etc.) may be outweighed by sufficiently high positive values or an extended period of low positive value. From the perspective of Negative Utilitarianism (NU) whose primary goal is to avoid disvalues this consequence is unacceptable; once a (dis-)value falls below a certain threshold, the doctrine of NU prescribes that the corresponding alternative must be avoided no matter what additional positive effects it might have. For example, according to NU a small amount of pleasure for a large group of people does not justify the otherwise avoidable death of one person. At the same time NU must allow for the ability to compare disvalues below a given threshold amongst each other, and so simple prescriptive rules like 'avoid deaths at any cost' do not suffice. If for instance all alternatives involve some quantifiable risk of death, a simple rule would recommend none of them whereas it is obvious that the alternative ought to be chosen that minimizes the death toll. We propose a layered approach to the evaluation of time-continuous action alternatives that reconciles the negative utilitarian doctrine with the EU approach. In our approach, an alternative A may be outweighed by an alternative B (written $B \succ A$) even if the overall sum of utility of A is larger than that of B. Formally, the account represents a mixture of EU and the well-known Maximin decision principle. We start by illustrating the use of a time-continuous value function in the non-layered case under certainty (Section II.A), then proceed to lay out the layered approach under certainty (Section II.B), address risk and uncertainty (Section III), and end with a brief philosophical discussion and a summary.

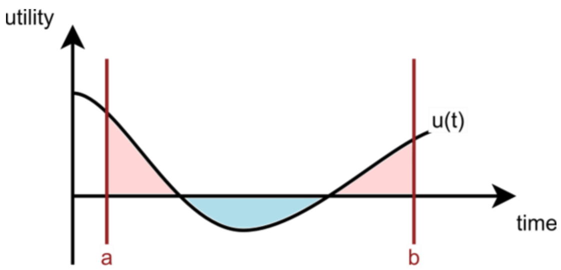

Figure 1. Summing up utility over time.

II. UTILITY OVER TIME UNDER CERTAINTY

A. *Summing up values and disvalues*

Let there be a finite set $\mathcal{A}$ of alternatives $A, B, C, \ldots$ under consideration and suppose we can determine units of utility by functions $u_x: \mathbb{R} \to \mathbb{R}$ from points in time to utility for each alternative x. Although it is common to assume that utilities are positive, this assumption is usually just made for technical convenience and since we wish to talk about negative utilitarianism we need to be able to distinguish value from disvalue. Hence, we assume that the range of utility functions is $[-1, 1]$. In this setting the total utility of an alternative x under certainty within closed time interval [a, b] is given by the definite integral

$$\int_a^b u_x(t)dt$$

Research on this article was conducted within the project PTDC/FIL-FIL/110117/2009 as well as partially funded by the individual research grant SFRH/BPD/84612/2012 of the Portuguese Foundation for Science and Technology (FCT), for which the author would like to express his gratitude.

Figure 1 depicts this case. Disvalues and values are summed up, in the continuous case by the definite integral, and may therefore outweigh each other arbitrarily. (In economics it is common to use a decay function to discount value over time but this and other refinements are left out for simplicity.)

B. A Layered Approach

The layered approach may be characterized by the following rules:

i. Tolerable small disvalues may be summed up with values.
ii. There are one or more negative layers $l_0 > \vartheta_1 \geq l_1 > \vartheta_2 \geq l_2 > \vartheta_3 \ldots \geq l_n > \vartheta_n \geq l_{n+1}$, where $\vartheta_1, \ldots, \vartheta_n$ are negative thresholds.
iii. Only alternatives within the lowest layer are taken into account.
iv. Within each layer alternatives can be compared by summing up their utilities within that layer over time.

Notice that in our terminology the lowest layer l_0 occurs highest in the graph; so the higher the layer, the less preferable it is. Figure 2 illustrates the approach: $A \succ B \succ C$ within $[a, b]$, because (i) A and B are in l_0 but C is in l_1, and (ii) the sum according to (1) of A is higher than that of B in this interval. Formally, the decision procedure works as follows. Let G be the set of alternatives and $L(x, a, b)$ denote the layer of x in the interval $[a, b]$, i.e. the layer in which the global minimum of $u_x(t)$ within the interval resides. We then compute the subset $G^*(a, b) := \{x | x \in \min_x L(x, a, b)\}$ of alternatives in the lowest layer. These are the only alternatives to consider. Let γ denote the upper threshold of the layer of the elements in this set if there is one, and otherwise $\gamma = 0$ for layer l_0. The elements of $G^*(a, b)$ are ordered as follows. If $\gamma > 0$, then

$$x \succcurlyeq y \Leftrightarrow \int_a^b (\gamma - u_x(t))dt \leq \int_a^b (\gamma - u_y(t))dt$$

and otherwise

$$x \succcurlyeq y \Leftrightarrow \int_a^b u_x(t)dt \geq \int_a^b u_y(t)dt$$

as before. The set of recommended alternatives is the maximum $\{x \mid x \succcurlyeq y\}$ for any $x, y \in G'(a, b)$ and the ordering relation $\succcurlyeq$ determined as laid out above. The corresponding strong preference and indifference relations are defined from this ordering as usual.

III. Risk and Uncertainty

Although we have also used it for value functions above, the term 'utility' should perhaps be used only for decision making under risk [3: 35-6; 5]. In a setting with risk the expected utility is $EU(x) = P(x)u(x)$ for an alternative x with probability $P(x)$. In our simplified time-continuous model, not taking into account decay, multiple attributes and changes in probability estimates over time, this translates to a linear

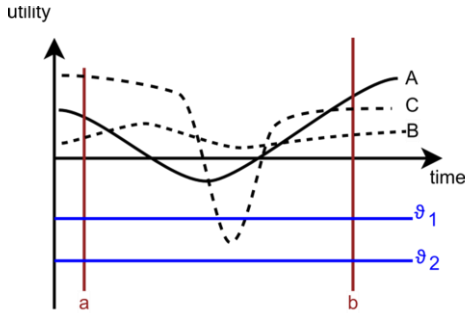

Figure 2. The Layered Approach.

transformation $EU(x, t) = P(x)u_x(t)$, i.e. a contraction of the curve. In this case the same decision procedure as laid out in the previous section may be applied.

In reality, however, risk can barely be determined exactly and epistemic uncertainty increases the farther an alternative extends into the future. One way to represent uncertainty of this kind is to associate with each alternative a pair of utility functions (u_x^+, u_x^-), each of which yields a maximally optimist and a maximally pessimist estimate. The functions will often be symmetric to each other around their mean and their difference is generally increasing with time, yielding a 'cone of uncertainty' representation as depicted in Figure 3. How ought one deal with this type of uncertainty in the present framework?

We believe that no general recommendation can be given for the layered approach, as this depends on the way the best and worst case scenarios have been estimated on one hand, and on the other hand from the perspective of NU also depends on what is at stake. Sometimes when the stakes are high it might make sense to remain cautious and use only the u^-estimates, thereby applying a version of the Precautionary Principle [6], whereas another time using the arithmetic mean $(u^+(t) + u^-(t))/2$ seems to be more adequate. We are not aware of any decisive criterion for deciding when to choose the former and when to choose the latter approach.

IV. Philosophical Discussion

A detailed discussion of the technical issues regarding our proposal would go beyond the scope of this article, and refinements such as multiple criteria and better representations of uncertainty are left for another occasion. It would also go beyond the scope of this article to address various known counter-arguments against NU such as the 'Pinbrick Argument' or Smart's argument [7] in response to Popper [8]. Let us, however, note that these arguments address cruder versions of NU than the one presented here, which correspond to strict rules, whereas we are able to compare alternatives within the same layer as usual. Something must be said about our implicit use of the Maximin principle, though. It is well known that using Maximin as a decision principle leads to paradoxes [9], because it discards too much information. The principle only takes into account worst case scenarios. By the same token, the layered approach discards potentially useful information; it applies Maximin to the choice of layers. Moreover, we have suggested that a cautious decision maker might sometimes use the lower boundary $u^-(.)$ for decision making under uncertainty. Hence, it seems reasonable to expect undesirable results for certain scenarios just like in applications of pure Maximin.

Apparent counter-examples may even be structurally identical to those motivating NE in the first place. Take for instance a decision in safety or health care with an alternative *A* involving a minimal, yet quantifiable that some person's might lose their life. Assume any alternative with potential death of few persons resides in layer *1*; by design, an alternative *B* in layer *0* would be recommended even in case the benefits of *A* might otherwise greatly outweigh the benefits of *B*.

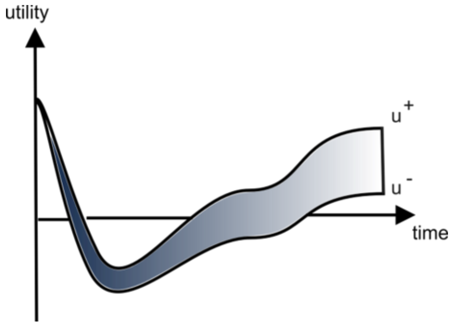

Figure 1. Summing up utility over time.

The problem is that *sometimes* we are willing to take such risks, for example when soldiers are supposed to die for 'a greater good', whereas in other cases like reactor safety we would not be willing to take them, and two such scenarios might be structurally identical. If we accept this as a premise, then it seems on one hand that negative utilitarianism lacks a criterion. Fixing the actual thresholds in the layered approach must be considered an open problem of moral philosophy. On the other hand, this problem is also indicative of our proposal's main advantage: The approach forces decision makers to make moral thresholds explicit which might otherwise remain implicit in the utility functions used, thereby committing them to a particular moral viewpoint and making them accountable for it. While the constraints encoded by layers could in principle also be expressed directly by a suitably chosen utility over time function, our approach deliberately separates utility from moral valuation by implicitly rejecting the preference-satisfaction view in the moral realm. If Broome is right by stating that "... the preference satisfaction theory is obviously false, and no one really believes it" [10: 4; cf. 11: 37-38], this ought not be a problem.

REFERENCES

[1] von Neumann, J. & Morgenstern, O. (1947), *Theory of Games and Economic Behavior*, Princeton UP, Princeton, NJ.

[2] Fishburn, P. C. (1970), *Utility Theory for Decision Making*, John Wiley and Sons, New York, London, Sidney, Toronto.

[3] Eisenführ, F.; Weber, M. & Langer, T. (2010), *Rational Decision Making*, Springer.

[4] Savage, L. J. (1954), *The Foundations of Statistics*, Wiley, New York.

[5] Schoemaker, Paul J. H., 'The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations', Journal of Economic Literature Vol. XX (June 1982), 529-563.

[6] Sandin, P. (2004), 'The Precautionary Principle and the Concept of Precaution', *Environmental Values* **13**, 461-75.

[7] Smart, R. N. (1958), 'Negative Utilitarianism', *Mind* **67**, 542-3.

[8] Popper, K. (1945), *The Open Society and Its Enemies*, Vol. 1, Routledge.

[9] Harsanyi, J. C. (1975), 'Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory', *The American Political Science Review* **69**(2), 594-606.

[10] Broome, J. (1999), *Ethics out of Economics*, Cambridge University Press, Cambridge, UK.

[11] Beckerman, W. (2011), *Economics as Applied Ethics: Value Judgements in Welfare Economics*, Palgrave MacMillian, New York.