Infinite, invariant, impartial consequentialism

Extended abstract

In decision theory, it is old news that infinity causes problems. We face a range of daunting cases in which either infinite expected rewards produce counterintuitive judgements (Pascal, 1670; Bernoulli, 1738), or in which infinite state spaces produce undefined totals which, in turn, render our judgements indeterminate (Nover & Hájek, 2004; Hájek & Nover, 2006; Colyvan, 2008).

Infinities sure are a pain. How lucky we are that we do not face such cases in ethics! Alas, it seems that we do, at least if we are consequentialists.

We live in an infinite world – our universe is spatially and temporally infinite, according to our best cosmology. We can also expect there to exist infinitely many instances of positive, finite value (by any plausible axiology) (see Garriga & Vilenkin, 2001; Knobe *et al.* 2006; Carroll, 2017). Given that this value does not consistently decrease in value or frequency, the total aggregated value in the universe will be infinite. And our present actions cannot change that fact.

If we are standard consequentialists – that is, if we aggregate value, weight all value equally, and seek to maximise the total – then, in every decision, we must maximise over a set of infinite values of equal magnitude. None of these infinite totals can be said to be greater than any other. Standard consequentialism, therefore, either cannot make any judgements at all, or it must permit everything (Nelson, 1991; Garcia & Nelson, 1994; Hamkins & Montero, 2000).

Neither universal silence nor universal permissiveness are attractive options for a moral theory. Consequentialism, then, is in a sorry state, and perhaps we had best reject it in favour of another moral theory. It is either that or it seems we must solve an intractable infinite decision problem.

Fortunately, various solutions have been proposed, each of which modify standard consequentialism to be able to judge among infinite outcomes, whilst departing from intuition as little as possible. These include the expansionist approach (Vallentyne, 1993; Vallentyne & Kagan, 1997; Arntzenius, 2014), the use of hyperreal numbers (Bostrom, 2011), bounded domains (Mulgan, 2002; Bostrom, 2011), limit-discounting (Jonsson & Voorneveld, forthcoming), and various others.

Foremost among these is expansionism. Rather than aggregating all value in the world and comparing the total, expansionism recommends that we aggregate local value in a particular order and compare the *cumulative* total. Those worlds whose totals increase more rapidly – those which overtake the alternative worlds and remain in the lead forevermore – are considered better. As for the order in which local value is aggregated, we would like this order to be both natural and essential. The only such order which solves the problem is the positioning of local value in space and time. (Note that each of the approaches mentioned above also require an order such as this to judge outcomes.)

Some consequentialists may immediately dislike this approach. It attributes moral relevance to location which, admittedly, is a departure from traditional consequentialist thought (Van Liedekerke, 1995; Bostrom, 2011). Indeed, expansionism has received little criticism beyond this. But, in evaluating any approach within infinite ethics, we must not demand the impossible. Specifically, it has been shown that it is impossible to construct any rule for ranking infinite worlds which: (i) is indifferent to (even finite) changes in the locations of value; (ii) satisfies even a weak dominance principle; and (iii) is both complete and transitive (Zame, 2007; Lauwers, 2010). In light of this, if we are to remain consequentialists, we must sacrifice the one principle to save the many. Indifference to locations is plausibly the least crucial of these and, at the least, we cannot dismiss expansionism purely on its violation (nor the many other approaches which make similar violations). In this paper, I will not ask the impossible – that expansionism satisfy all of the above conditions. Instead, I will show that it cannot concurrently satisfy completeness, transitivity, and the requirement that moral judgements be absolute.

According to special relativity, positioning in space and time is not absolute. It varies with the speed of the observer (Einstein, 1905; Poincaré, 1905). This is problematic for a moral theory which relies on

spatiotemporal position to make judgements. In this paper, I construct scenarios in which expansionism provides conflicting judgements at different speeds. This demonstrates that, under expansionism, moral judgements are relative, to the agent or observer's speed. This 'moral relativism' appears absurd, and plausibly worse than rejecting consequentialism entirely.

Luckily, there are two possible escapes for the expansionist. They may choose to (i) remain silent in cases of conflicting judgements, or (ii) be indifferent between options in those cases. The former option entails an alarming degree of incompleteness (and one which cannot be solved by further strengthening, as it arises from the theory *already* being too strong). The latter option entails intransitivity of indifference, and hence a susceptibility to money pumps and Dutch Book scenarios. Both options are implausible, as is the alternative of accepting relativism. Therefore, expansionism cannot be accepted (nor, it should be noted, can the many other approaches which are also reliant on spatiotemporal order and suffer a similar fate).

I propose a solution to the relativistic problem. The problem arises partly due to an underlying assumption – that value is instantiated at discrete locations, which correspond to points in spacetime. To see why this is the case, consider the *spacetime interval*, the one measure of spatiotemporal distance (by which we might order locations) which does not vary with speed. The reason we cannot simply use it to order locations is that each value of the spacetime interval identifies an infinite number of points. If value occurs at individual points, then this may give us infinite value for every single one of our cumulative totals. We face the same problem as we did to begin with.

Fortunately, however, for any value of the spacetime interval s, the region of spacetime with distance less than s has only finite volume. We can ensure that our cumulative sum remains finite if we locate value not at discrete points but at *regions* of continuous points. We can take value as being proportional to the volume of the region. (This will correspond to, for instance, the duration of a phenomenal experience.) Then, we will only be able to fit finite value into each step of our aggregation, and into each of our cumulative totals. We can then use the spacetime interval after all, and we have an order which is speed-invariant and gives finite cumulative totals.

By making these modifications, we can give a speed-invariant, impartial method of judging among infinite outcomes. (A similar solution can also be applied to the other approaches.) This will not be entirely complete, as expansionism (and other approaches) are still incomplete in some non-relativistic cases. Nonetheless, we have made progress.

It is also worth noting that this solution to the relativistic problem has substantive implications, so we might have made more progress than we thought. My solution implies (and, indeed, requires) that value be proportional to the volume of spacetime it occupies. This confirms Bentham's (1781: 21) claim that value be proportional to duration. It also lends some support to the hedonist and welfarist views – after all, for their rivals, such as preference-satisfaction consequentialism, it is difficult to assign and to evaluate the duration of valuable phenomena. In addition, for the dilemmas raised by Cain (1995; 2005) and Almeida (2010), my solution produces determinate judgements.

Bibliography

Almeida, M. (2010). 'Two Challenges to Moral Nihilism.' *The Monist* 93, no. 1: 96-105.

Arntzenius, F. (2014). 'Utilitarianism, decision theory and eternity', *Philosophical Perspectives*, 28/1: 31-58.

Bentham, J. (2000) [1781]. *An Introduction to the Principles of Morals and Legislation*. Kitchener: Batoche Books.

Bernoulli, D. (1954) [1738]. 'Exposition of a new theory on the measurement of risk.' *Econometrica* 22: 23-36.

Bostrom, N. (2011). 'Infinite ethics', *Analysis and Metaphysics*, 10/1: 9-59.

Cain, J. (1995). 'Infinite utility.' *Australasian Journal of Philosophy* (1995): 401-404.

Cain, J. (2005). 'Utilitarianism and the moral significance of an individual.' *Southwest Philosophy Review* 21, no. 1: 53-60.

Garcia, J. L. A., and M.T. Nelson. (1994). 'The problem of endless joy: is infinite utility too much for utilitarianism?.' *Utilitas* 6, no. 2: 183-192.

Carroll, S.M. (2017). 'Why Boltzmann Brains Are Bad.' *arXiv preprint* arXiv:1702.00850 (2017).

Colyvan, M. (2008). 'Relative expectation theory.' *The Journal of Philosophy* 105, no. 1: 37-44.

Einstein, A. (1905). 'Zur elektrodynamik bewegter körper', *Annalen der physic*, 322/10:891-921.

Hájek, A., and H. Nover. (2006). 'Perplexing expectations.' *Mind* 115, no. 459: 703-720.

Hamkins, J.D., and B. Montero. (2000). 'Utilitarianism in infinite worlds.' *Utilitas* 12, no. 1: 91-96.

Garriga, J., and A. Vilenkin. (2001). 'Many worlds in one.' *Physical Review* D 64, no. 4: 043511.

Jonsson, A., and M. Voorneveld. (forthcoming). 'The limit of discounted utilitarianism.' *Theoretical Economics*.

Knobe, J., K.D. Olum, and A. Vilenkin (2006). 'Philosophical implications of inflationary cosmology', *The British Journal for the Philosophy of Science,* 57/1: 47-67.

Lauwers, L. (2010). 'Ordering infinite utility streams comes at the cost of a non-Ramsey set', *Journal of Mathematical Economics*, 46/1: 32-7.

Mulgan, T. (2002). 'Transcending the infinite utility debate', *Australasian Journal of Philosophy* 80, no. 2: 164-77.

Nelson, M. (1991). 'Utilitarian eschatology', *American Philosophical Quarterly*, 28/4:339-47.

Nover, H., and A. Hájek. (2004). 'Vexing expectations.' *Mind* 113, no. 450: 237-249.

Pascal, B. (1670). *Pensées*. Dezobry et E. Magdeleine.

Poincaré, H. (1905). 'Sur la dynamique de l'électron', *Comptes Rendus*, 140/1:1504-8.

Smith, Q. (2003). 'Moral Realism and Infinite Spacetime Imply Moral Nihilism', in H. Dyke (ed.) *Time and Ethics: Essays at the Intersection*, 43-54. Berlin: Kluwer Academic Publishers.

Vallentyne, P. (1993). 'Utilitarianism and infinite utility', *Australasian Journal of Philosophy*, 71/2:212-7.

Vallentyne, P. and S. Kagan (1997). 'Infinite value and finitely additive value theory', *Journal of Philosophy*, 94/1:5-26.

Van Liedekerke, L. (1995). 'Should utilitarians be cautious about an infinite future?.' *Australasian Journal of Philosophy*: 405-407.

Zame, W. (2007). 'Can intergenerational equity be operationalized?', *Theoretical Economics*, 2/2.