Journal of Philosophy of Life Vol.2, No.1 (March 2012):55-66
[Discussion Paper]

Negative Average Preference Utilitarianism

Roger Chao*

Abstract

For many philosophers working in the area of Population Ethics, it seems that either they have to confront the Repugnant Conclusion (where they are forced to the conclusion of creating massive amounts of lives barely worth living), or they have to confront the Non-Identity Problem (where no one is seemingly harmed as their existence is dependent on the "harmful" event that took place). To them it seems there is no escape, they either have to face one problem or the other. However, there is a way around this, allowing us to escape the Repugnant Conclusion, by using what I will call Negative Average Preference Utilitarianism (NAPU) – which though similar to anti-frustrationism, has some important differences in practice. Current "positive" forms of utilitarianism have struggled to deal with the Repugnant Conclusion, as their theory actually entails this conclusion; however, it seems that a form of Negative Average Preference Utilitarianism (NAPU) easily escapes this dilemma (it never even arises within it).

1. Introduction

For many philosophers working in the area of Population Ethics, it seems that either they have to confront the Repugnant Conclusion, or they have to confront the Non-Identity Problem. To them it seems there is no escape, they either have to face one problem or the other. What I will try to show in this paper however, is that there is a way around this, allowing us to escape the Repugnant Conclusion, by using what I will call Negative Average Preference Utilitarianism (NAPU) – which though similar to anti-frustrationism, has some important differences in practice.

First, let us look at how Derek Parfit characterizes the Non-Identity Problem:

> "There are two rare conditions, J and K, which cannot be detected without special tests. If a pregnant woman has Condition J, this will cause the child she is carrying to have a certain handicap. A simple treatment would

* Curtin University, Perth 6102, Western Australia, Australia. The Author can be contacted at rogersteppe[a]gmail.com.

> prevent this effect. If a woman has Condition K when she conceives a child, this will cause this child to have the same particular handicap. Condition K cannot be treated, but always disappears within two months. Suppose next that we have planned two medical programmes, but there are funds for only one; so one must be cancelled. In the first programme, millions of women would be tested during pregnancy. Those found to have Condition J would be treated. In the second programme, millions of women would be tested when they intend to try to become pregnant. Those found to have Condition K would be warned to postpone conception for at least two months, after which this incurable condition will have disappeared. Suppose finally that we can predict that these two programmes would achieve [the same] results in as many cases. If there is Pregnancy Testing, 1,000 children a year would be born normal rather than handicapped. If there is Preconception Testing, there would each year be born 1,000 normal children rather than a 1,000, different, handicapped children."[1]

Now given this, is there a morally justifiable reason to cancel one program over the other? Most people would feel that both programs are equally valuable as they have the equivalent effects on the parents, and both programs result in 1000 normal children being born instead of 1000 handicapped children. However if the testing for Condition J is cancelled, 1000 handicapped children would be born, who would otherwise not be handicapped, thus cancellation of testing for Condition J would be worse for those children. If testing for Condition K is cancelled however, 1000 children would also be born handicapped, however these children would never have existed if there was testing for Condition K, and therefore the cancellation of testing for Condition K cannot be said to be worse for them.

To escape this non-identity problem, our only recourse is to say that both programs are equal and that it makes no moral difference which one is cancelled, as either way 1000 normal children being born instead of 1000 handicapped children. However, what this leads to then, is the principle that "if other things are equal, it is better if there is a greater sum total of happiness."[2] However when we apply this to populations, it results in the Repugnant Conclusion.

---

[1] Derek Parfit, *Reasons and Persons*, p.136.
[2] Ibid.

A theory of population ethics necessitates arriving at the Repugnant Conclusion if and only if for a world W, with population of size X, which has an individual utility level U (where U>0), there is another possible world W*, with population size X* (where X*>X), which has an individual utility level U* (where U*<U), W* is preferred to W. In general terms, the Repugnant Conclusion is implied whenever increases in X can be substituted for decreases in U, no matter how close to zero U (as long as it is positive) gets.

2. Negative Utilitarianism – Not So Easily Discredited

Current "positive" forms of utilitarianism have struggled to deal with this dilemma, as their theory actually entails this conclusion; however, it seems that a form of Negative Average Preference Utilitarianism (NAPU) easily escapes this dilemma (it never even arises within it).

Normally the criticism thrown at negative utilitarianism is that the best way to minimize suffering is to destroy the whole world.[3] With no sentient life, suffering is thus minimized.[4] This is generally used as the knockdown argument to discredit negative utilitarianism. However, because of such widespread condemnation of negative utilitarianism as a result of this, very little work has been done to show how other forms of negative utilitarianism are not defeated by this argument, and can solve the Repugnant Conclusion. It seems that negative utilitarianism has "had its day" and has just been left on the wayside, as another of those outdated, defeated theories. What I will attempt to argue in this paper however, is that this is not actually the case, and we should thus not reject it as quickly as many people do.

There is a way out from this extremely counter intuitive conclusion, if we look carefully at what is known as the experience requirement. For a classical utilitarian, if someone "harms" you but you never experience it, they have not done anything wrong (and thus it is not really a "harm"), since utility consists of pleasurable mental states. Imagine for example, that one of your friends purposely spreads vicious rumours about you behind your back for your whole life (which you never ever find out); from which you experience no ill effects (it

[3] Smart, J.J.C, & Williams, B, *Utilitarianism, For and Against*, p.29.
[4] Griffin, James, 'Is Unhappiness Morally More Important Than Happiness?,' *The Philosophical Quarterly*, Vol. 29, No. 114, p.48.

does not change how other people in the world treat you). Thus, seemingly it does not hurt you and thus is not wrong, as what you do not experience, cannot hurt you. This is often called “the experience requirement”; hedonism requires that the agent experience something in order for that thing to be good (or bad) for them. You never experience any “displeasure” from this and thus to the classical utilitarian (hedonist) are not harmed by it, since they accept this experience requirement.

To the preference utilitarian however, this is wrong, as even though you do not experience any ill effects as a result of this, you still have a preference for a good reputation, a preference for your friends to speak well of you, a preference to have friends you can trust etc. Furthermore, even if the whole world were to be destroyed instantly, people’s preferences to live would still be violated, for even if they do not experience pain when the world is destroyed their preference to keep on living is frustrated. Thus it seems that preference utilitarianism can reject this experience requirement, to show that even though you never experience direct harm from another agent’s actions, their actions can still be wrong as you have a preference for this harm not to be done, which is frustrated.

Now the objection to negative utilitarianism about “destroying the world to abolish suffering,” only really works if we are looking at a classical form of negative utilitarianism. However, if we look at what I will call “negative preference utilitarianism” this objection no longer holds any weight. As we have seen, a classical utilitarian can only appeal to indirect wrongness (how it affects other people) in painless killing and not direct wrongness, as the person killed is not harmed themselves. They thus struggle with why it is wrong to painlessly kill a hermit who no one even knows exists, let alone will miss (since no other people are affected). In parallel to this, negative utilitarianism also can only appeal to direct wrongness in a case of painless killing such as instantly detonating 100 nuclear bombs on earth to kill all its inhabitants. It thus struggles to see why this is wrong too, since if all the earth’s inhabitants were to be killed, no one would be around to suffer because of this.

In the case of the hermit, this problem is readily solvable for a preference utilitarian, where it does not matter that he is a hermit, or that he is killed painlessly. Rather what matters here, is that his preference for continued life is frustrated, and thus that is why it is wrong to kill him, even though neither he nor anyone else suffers. That is why to the classical utilitarian, there is nothing wrong with killing the hermit, yet to the preference utilitarian there is.

Now using this analogy, we can now look at it from a negative preference utilitarian viewpoint, and see that the same can be said about destroying the world. As once you exist, you (and the other people who know you) all have preferences concerning your existence (their preference to go on holiday with you, your preference to live, their preference to have you as a friend, your preference to go to the concert next week etc.). Thus, if you were to be killed, all these preferences would be frustrated, and would be prima facie wrong. The majority of sentient beings in the world have a very strong preference for continued existence, thus killing them violates this and is prima facie wrong, even if none of them directly suffers as a result. Thus, we can see that by appealing to a form of "negative preference utilitarianism" as opposed to classical negative utilitarianism, we can easily escape the objection. Thus, it seems that the negative preference utilitarian is not committed to destroying the whole world, but rather has to take into consideration that there is a cost to destroying the world, which needs to be weighed against the cost involved in continued existence.

Another objection many people have sought to use to discredit and debunk negative utilitarianism is by saying that to a negative utilitarian, it would be better if the world had never existed in the first place — even if the only painful experience that would ever exist was a small paper cut. This objection tries to play on people's intuitions that if the only suffering that would ever exist in the world was a sole paper cut, that this is still better than the world never existing, which they claim is counter to what the negative utilitarian believes. However, this argument is false as it is based upon an impossible assumption. This argument contains an implied premise (which is necessary, given that someone is giving this argument) that there is a conscious being evaluating existence without existing themselves, which is necessarily impossible.

The actual fact however, is that if the world had never existed, there would be no regret, longing, or guilt about not existing for this evaluating agent to experience, (as there would be no one to experience this); thus there would be no one to make this argument. In this thought experiment, we find it counter-intuitive (we feel some "loss") if the world were to not exist, because we are the agent doing the evaluating, and we currently exist, and thus prefer this to not existing. However, this would not actually be the case if the world had never existed, as there would no one around to experience this "loss" in the first place (or in hindsight). Thus, this argument is fallacious due to its

impossible implied premise.

3. Breaking the Argument Down

There is a huge difference (they are very different questions) as to:

1. Whether it is preferable to be brought into existence or not in the first place

And

2. Once you already exist, if it is better to then be brought out of existence.

Above I have shown how negative preference utilitarianism is not defeated by the usual "destruction of the world" objection, and thus supports the negative answer to question 2. However, in order to avoid the Repugnant Conclusion, we must then have a way to support the affirmation of question 1, although only in so far as it does not commit us to the Repugnant Conclusion.

Thus as I have already shown, Negative Preference Utilitarianism isn't as easily defeated as most people think, but what we must now show, is why it is sometimes preferable to not be brought into existence in the first place.

When it comes to utilitarianism, there is the total and the average view. Now there are also two perspectives from which we can look at this, one perspective is that where we are deciding what to do with presently existing people, the other perspective is that where we are deciding whether to bring more people into existence.

With the total view, where the total amount of happiness is all that matters, we are seemingly committed to the mere addition paradox (the Repugnant Conclusion) by maximizing total happiness.[5] Thus, it seems that we must then look to the average view. The average view tries to maximize the average (per capita) utility. However, this seemingly implies that a population with 1 million members, (500,000 with 100 utils, and 500,000 with 99 utils) is worse than a population of 500,000 people (each with 100 utils), leading to a reverse of the Repugnant Conclusion. It seems to imply that for any fixed population, what they would be morally required to do is kill off those members who have the least utility, to raise the average utility level. Thus, it seems that classical

[5] Cowen, T, 'Resolving the Repugnant Conclusion', in Ryberg, J & Tännsjö, T (Eds.), *The Repugnant Conclusion: Essays on Population Ethics*, p.81.

utilitarianism commits us to both these highly counter intuitive conclusions.

However if we look at what classical negative utilitarianism requires of us, it seems just as counter intuitive as what classical utilitarianism requires of us. In its total form it requires that we reduce total suffering (thus meaning we should kill as many people as possible, leading to a reverse of the Repugnant Conclusion (where no one exists) again). In its average form, it requires us to reduce the amount of suffering per capita, thus leading us to kill off the people who suffer most.[6] Thus classical negative utilitarianism does not have a way out of this either, and thus classical utilitarianism, and classical negative utilitarianism are both fatally flawed.

The next thing we can look at is (positive) preference utilitarianism. In its total form however, it again forces us to accept the Repugnant Conclusion (maximizing the total preferences satisfied, in terms of quality and quantity). In its average form it still fails as it seems to force us to reject the principle of mere addition. This principle states, "the addition of extra worthwhile lives which do not affect anyone else, cannot make the outcome worse," which seems quite (very) in line with our moral intuitions. However in terms of average preference utilitarianism, in a population of 100 people, each with 100 utils, it is "wrong" to bring one more person into existence with a level of 99 utils (still an extremely high level, well above the minimally decent life). This is so even though the rest of the population will not be affected (as it reduces the average level of preference satisfaction). Thus again it seems we must reject preference utilitarianism in its positive form as highly counter intuitive, as what it necessitate as "wrong" is highly counter intuitive since it violates "mere addition."

However if we look to negative preference utilitarianism, this dilemma is readily solvable from an average viewpoint, and this is what I aim to show in the rest of this paper. The total viewpoint still results in counter intuitive conclusions; however, the average viewpoint does not.

On the total view, we will first look at the perspective of what we should do with presently existing people. With the aim of the total view being to minimize total preference frustrations, it seems that we cannot just go around killing people, as since they are living, they must prima facie have a strong preference (relatively) for continued life (even if you/they aggregate all their past

[6] Ibid.

preference frustrations) for that is why they haven't (ceteris paribus) killed themselves. Thus killing them will create even more (in terms of quality/strength) preference frustrations — preferences to go to the beach tomorrow, to drink wine on Sunday etc.

Now if we then look at what to do about bringing new people into existence, the total view would say we should not bring any new people into existence (unless it can be guaranteed they would not experience any preference frustrations). However since this is near impossible, it thus seems it commits us to not bring any new people into existence, thus seemingly disallowing us from having children. This however is highly counter intuitive, and thus it seems we must reject this total viewpoint for negative preference utilitarianism.

Let us now look at this from an average perspective, with presently existing people (with preferences for continued existence). Now if our aim is to reduce preference frustration per capita (in both quality and quantity), we cannot just go around killing the people with the highest level of preference frustration. This is because ceteris paribus they will most likely have a very strong preference to live (or they would have killed themselves), and preferences for other goals in life that would be frustrated if we kill them (a preference to live/survive is ceteris paribus probably your strongest preference). Thus from an average perspective (negative preference utilitarianism) when looking at presently existing people, we should ceteris paribus not kill them (which sits perfectly in line with our moral intuitions).

Let us now look at it from the perspective of deciding whether to bring more people into existence. From the average viewpoint, it seems that we can, as long as their expected level of preference frustration is lower than the average. Thus it does not forbid us from having children (which again fits perfectly with our moral intuitions), and it does not commit us to the Repugnant Conclusion, as we are allowed to have children, but we cannot sacrifice their quality (in terms of preference frustration) for quantity (more children).

Thus, it seems that the average negative preference utilitarian viewpoint is the most intuitively plausible. It does not result in any of the counter intuitive conclusions that classical negative utilitarianism does, it does not result in any new counter intuitive implications, and it does not force us to being committed to the Repugnant Conclusion. Thus, it seems that being an average negative preference utilitarian seems the best way out of the repugnant conclusion from a utilitarian standpoint.

4. Negative Average Preference Utilitarianism – A Summary

As we have seen, classical utilitarianism in its total form commits us to the Repugnant Conclusion, and in its average form, it commits us to killing the worst off, and thus both forms of classical utilitarianism have to be rejected. Preference Utilitarianism in its total form, also commits us to the Repugnant Conclusion, in its average form it violates the principle of mere addition (where the addition of extra worthwhile lives which do not affect anyone else, cannot make the outcome worse), and thus Preference Utilitarianism also has to be rejected. Classical negative utilitarianism in its total form, leads us to killing as many people as possible (the reverse Repugnant Conclusion) and in its average form leads us to killing the worst off, so classical negative utilitarianism has to be rejected. Thus, all that remained left was negative preference utilitarianism. In its total form however it made having children impermissible (and was thus highly counter intuitive), and thus a total approach had to be rejected as well.

The final remaining option was the average viewpoint. It allows us to have children, yet avoids the Repugnant Conclusion, does not allow us to just kill people, and has no counter intuitive implications. Thus, overall it seems that the best/only way out of this Repugnant Conclusion, which is still in line with our moral intuitions is that of a negative average preference utilitarian (NAPU).

NAPU does not commit us to the Repugnant Conclusion, as the Repugnant Conclusion is all about increasing the average preference frustration level, which NAPU explicitly forbids. The Repugnant Conclusions is all about increasing the quantity of existing people, to exceed the loss in quality of their lives (as long as they are above the level of neutral existence). That is why the Repugnant Conclusion results in extremely large numbers of people living with an extremely low quality of life (all be it, above the minimally decent standard of living). NAPU on the other hand is about reducing the average level of preference frustrations, and thus vehemently denies the Repugnant Conclusion. NAPU concerns itself solely with necessarily existing persons, those who currently exist or will necessarily exist by our actions. In deciding whether to bring new people into existence (people who will not necessarily exist), we must remember that if we do not bring them into existence, that they will not have

preferences,[7] and thus that they do not currently have preferences to exist.

Thus in NAPU it should be allowed (or at least not blamed) that a person with the worst utilities is left to suffer, whilst other people's utilities are improved because this increases the average. Some people might argue that this goes against Rawls' Difference Principle, in that NAPU only thinks about the average and not a specific person and thus means that in some instances the most miserable person should not be saved. However, NAPU is a form of triage; it is about maximizing resource usage to gain the most efficient outcome. Rawls would say that amongst two people, with extremely similar levels of pain differing only by a minuscule amount, reducing the worst off person's pain by 0.00001% is better than reducing the next worst off person's pain by 99%. NAPU however, pragmatically realizes that with finite resources, these resources should be spent on the person for whom the greatest benefit is realized.

5. Comparison to and Objections from Other Theories, and Its Relevance to Philosophy of Life

Thus, we have already seen how intuitively plausible negative average preference utilitarianism is, and how it escapes the Repugnant Conclusion without any counter intuitive implications. However, on the surface, it looks very similar to Fehige's anti-frustrationism. There are however, some fundamental differences that set it aside, as a superior ethical theory.

Anti-frustrationism implies that having a preference satisfied just brings you back to the neutral level of welfare that you would have had anyway, if the preference had not even arisen.[8] Thus to the anti-frustrationist, because non-existing persons cannot have preferences, (and thus cannot have any frustrated preferences) this places them on par with someone who exists with absolutely all of their preferences satisfied instantaneously. Thus to the anti-frustrationist, those who do not exist, are always at least as well off than those who do exist, because most (or in reality all) existing persons have at least one frustrated preference, and thus have a lower level of wellbeing than a

[7] Benatar, D, 'Why It is Better Never to Come into Existence,' *American Philosophical Quarterly*, Vol. 34, No.3, p.345.
[8] Fehige, C, "A Pareto Principle for Possible People," in Fehige, C & Wessels, U, (Eds.) *Preferences*, p.511.

non-existing person.[9] Thus, it seems that the anti-frustrationist is committed to anti-natalism, something that most people find extremely counter-intuitive.

If we recall however, NAPU does not commit us to being anti-natalists. All NAPU says is that if the child you are going to have will have above average preference frustrations, then we should not have it. It does not ban us from having children, but rather only requires that we do not have children if doing so will result in a higher average level of preference frustration in the world. This is the difference between anti-frustrationism and NAPU, and is thus why NAPU is more intuitively plausible.

One objection might be that in the future, an extremely happy population is deciding whether to have children, but see that having a baby will result in a child who exceeds the average preference frustration level by a miniscule amount, and thus are prohibited from having it. Some people might find being banned from having children quite objectionable, and thus use this to object to NAPU. However, according to NAPU, if many parents really have an extremely strong preference to have children (which will be frustrated if they cannot have children), then their preference frustrations from this needs to be weighed up against it. Thus it may be the case that in some instances, if many parents have such a strong preference to have children, that they should be allowed to. For to not let them have children, will result in a higher average level of preference frustration, than allowing them to have the child (who will have an above average level of preference frustration) in the first place.

Thus as we have seen, utilitarianism comes in two forms, average or total, and these two forms can further be split into a hedonist or a preferentialist version. These four forms can then be further split into a negative and a positive version. Out of these eight versions it seems seven of them fail, whilst one has been overlooked – Negative Average Preference Utilitarianism.

Thus, what I hope to have shown in this paper is that negative utilitarianism cannot just be rejected as quickly as most people do. A form of this which I have called Negative Average Preference Utilitarianism (NAPU), seems to be the most intuitively plausible and appealing form of utilitarianism which solves the Repugnant Conclusion, without leading to any counter intuitive dilemmas. Whilst on the surface is looks very similar to anti-frustrationism, it does not result in the highly counter intuitive anti-natalism that anti frustrationism

[9] Ibid, p.513.

commits us to. Thus, I contend that Negative Average Preference Utilitarianism is a viable option to escape the Repugnant Conclusion from a utilitarian perspective.

This has implications for many facets in philosophy of life, ranging from sanctity of life arguments (where killing someone is only justified if it reduces the average level of preference frustration in the world, taking into consideration the preferences of the person being killed); to the value of human existence (to do with not having preferences frustrated); to the dignity of human life (the dignity of human preferences is the foundation of a moral vision of society) and many other fields, and is thus worthy of further exploration.

Bibliography

Benatar, D, "Why It is Better Never to Come into Existence," *American Philosophical Quarterly*, Vol 34, No.3, July 1997: pp. 345-355.

Cowen, T, "Resolving the Repugnant Conclusion," in Ryberg, J & Tännsjö, T (Eds.), *The Repugnant Conclusion: Essays on Population Ethics*, Boston, Kluwer Academic Publishers, 2004: pp. 81-97.

Fehige, C, "A Pareto Principle for Possible People," in Fehige, C & Wessels, U, (Eds.) *Preferences*, New York: W. de Gruyter, 1998: pp. 509-543.

Griffin, J, 'Is Unhappiness Morally More Important Than Happiness?,' *The Philosophical Quarterly*, Vol 29, No. 114, January 1979: pp. 47-55.

Parfit, D., *Reasons and Persons*, Clarendon Press, Oxford, 1984.

Smart, J.J.C, and Williams, B, *Utilitarianism, For and Against*, Cambridgeshire: Cambridge University Press, 1973.