<u>Draft</u> **Taiwan 2: Nov 6; Jan 28 2006**

<u>I have not yet written acknowledgement in my paper yet.</u>

CAN MANY HEADACHES MORALLY OUTWEIGH A HUMAN LIFE?

Hon-Lam Li

Department of Philosophy

Chinese University of Hong Kong

Shatin, N.T.

Hong Kong

Email address: **honlamli@hotmail.com**

Abstract: Can many minor headaches, each suffered by a different individual, morally outweigh a human being's life? For most people, an affirmative answer to this question would be counter-intuitive and morally repugnant, and hence must be false. However, utilitarianism would imply such a repugnant view, because it assumes that the utilities are interpersonally aggregable. That is, utilitarianism assumes that smaller utilities – even trivial ones – dispersed among many individuals can be summed up to outweigh morally a human being's life. But this utilitarian reasoning would beg the question against those who do not already accept the assumption that utilities are interpersonally aggregable. Despite this, it has been argued that a non-utilitarian starting point would eventually lead back to the repugnant view. One compelling argument that has been put forward consists of the following three premises:

(1) Between saving either an innocent person's life, or a number of persons each of whom would otherwise suffer near-death (a condition which is nearly as bad as death), we should save the latter, if the number is sufficiently high. The grounds for this is that one case of harm is always tradable with a greater number (n) of *slightly* lesser harms, and can always be outweighed by an even greater number (n+1) of *slightly* lesser harms.

(2) All sorts of harms form a continuous spectrum, such that starting from death, there is always a slightly lesser harm.

(3) If A morally outweighs B, and B morally outweighs C, and C morally outweighs D, …, and Y outweighs Z, then A morally outweighs Z.

Given these three premises, each of which seems highly plausible and perhaps even self-evident, it follows that a person's life can be morally outweighed by a large

number of headaches, each suffered by a different individual.

In this paper, I show that this argument is not good, because (3) is false. I argue that in comparing 1 case of death with 100 cases of near-death, approximation is involved. Such a comparison is one in two variables, namely, the seriousness of harm (h), and the number of people suffering it (n): Since near-death is so close to death, and since 100 far exceeds 1, we can presume, by means of making the approximation that near-death is (nearly) death, that 100 cases of near-death morally outweighs 1 case of death. We may wish to continue the process yielding the result that 100 cases of near-death morally outweigh 10,000 cases of near-near-death. By transitivity, 10,000 cases of near-near-death would morally outweigh 1 case of death. This kind of comparison cannot go on forever, however, because approximation is involved in each comparison, and approximation is not transitive. Thus, I argue that the argument that a huge number of headaches morally outweigh one case of death is blocked.

I also argue that the two-variable comparison of harms (as opposed to one-variable comparisons), when made repeatedly, is a Sorites Paradox. A Sorites Paradox has the following characteristics: (1) the values or properties of the two poles of a spectrum through which a Sorites Paradox traverses are opposite (e.g, there is a heap in one end, but not in the other end; someone is bald in one case, but not the other); (2) a Sorites Paradox involves many steps; (3) in each step, approximation is involved; (4) somewhere between the two poles there exists a range with vague or indeterminate boundary, and the value or property over this range is itself unclear or indeterminate; (5) there is something vague or indeterminate about the central concept or feature which the Sorites Paradox argument exploits throughout the steps. All of these characteristics are present in the current problem.

> I believe one should trust problems over solutions, … and pluralist discord over systematic harmony. Simplicity and elegance are never reasons to think that a philosophical theory is true: on the contrary, they are usually grounds for thinking it false.... Often the problem has to be reformulated, because an adequate answer to the original formulation fails to make the *sense* of the problem disappear. It is always reasonable in philosophy to have great respect for the intuitive sense of an unsolved problem, because in philosophy our methods are always themselves in question, and this is one way of being prepared to abandon them at any point. (Thomas Nagel, *Mortal Questions*, x-xi)

I. The Problem

Do many headaches, each had by a different individual, morally outweigh a person's life? If you can save either the life of an innocent person, or a large group of people each from having a headache, but not both, should you save the former or the latter? I believe that most people believe that we should save the innocent person. In fact, our intuition to save the innocent person's life instead of the group from the headaches is so strong that any view that requires us to do otherwise would seem morally repugnant and hence must be wrong.

Yet, utilitarianism would have us reason differently: Given that death and a headache both have disutilities, and that smaller utilities (or disutilities) among different people can be summed up – or are aggregable – then utilitarianism would lead to the conclusion that there exists a certain number, n, of headaches, such that the disutilities of these headaches would outweigh the disutility of the loss of a human life (unless a human life has infinite utility).

That lesser utilities dispersed among different individuals, no matter how trivial these utilities are, can be interpersonally aggregated to morally outweigh a much weightier claim someone possesses, provided the number of these trivial claims is large enough, is an unargued basic axiom of utilitarianism. Call this axiom the <u>Aggregability Axiom</u>. Unless this Axiom is correct – that is, that utilities are

---

[2] I am not against the view that intrapersonal utilities can be aggregated to outweigh a weightier claim. Thus if faced with (1) the prospect of having a huge number of headaches in one's life, and (2) the prospect of death, one can rationally choose death.

interpersonally[2] aggregable – it would make no sense to say that utilities can be maximized among a group of people. Thus, a utilitarian must accept the idea (1) that utilities are interpersonally aggregable, and hence (2) that a certain number of headaches would outweigh a human life unless a human life has infinite utility.

But it seems clear that a human life does not have infinite utility, unless it has infinite duration. For a life consists of finite duration of time: unless each minute has infinite utility, a life of finite duration cannot have infinite utility. But one minute of a life does not have infinite utility, or otherwise we would have wasted infinite utility waiting in a queue – which would be very counter-intuitive.

So utilitarianism has the consequence that even if a life has extremely high utility, there exists a sufficiently big number, n, such that n headaches can be interpersonally aggregated to outweigh a human life. Call this the Repugnant Consequence. Utilitarians would have to swallow such counter-intuitive consequence. But they argue that our intuitions are sometimes unreliable.

Now what about those who do not yet share the Aggregability Axiom?[4] They would not be persuaded to accept the Repugnant Consequence. Instead, they

---

[4] The Aggregability Axiom is the axiom that lesser utilities dispersed among different individuals, no matter how trivial these utilities are, are interpersonally aggregable to morally outweigh a much weightier claim someone possesses, provided the number of these trivial claims is large enough

[6] They include Alastair Norcross, Francis Kamm, Derek Parfit, as well as Sophia Reibetanz. See, for example, Reibetanz, "Contractualism and Aggregation," *Ethics*, vol. 108 (January 1998) 296-311.

would argue that utilitarianism can be refuted *reductio ad absurdum*, since given the utilitarian view that utilities are interpersonally aggregable, the Repugnant Consequence (which to them is clearly false) is inescapable.

However, it has been argued (by utilitariansas well as some non-utilitarians[6]) that even if the Aggregability Axiom is not presupposed, the Repugnant Consequence is still inevitable. Call this view <u>the Inevitability View</u>: If the Repugnant Consequence is inevitable from starting points that do not presuppose the Aggregability Axiom – that is, from non-utilitarian starting points -- then the Repugnant Consequence must be correct, after all. Consequently, the intuition that the Repugnant Consequence is repugnant may come to have little force.

I shall mainly consider one argument for the Inevitability View, an argument which seems to be the most forceful of all. (I shall consider another argument – briefly – toward the end of this paper.) It consists of three premises:

(1) **<u>Tradability Thesis</u>**: One case of harm is always tradable with a greater number (n) of *slightly* lesser harms. (The qualification of "slightly" is important here.) Indeed, **one case of harm can always be outweighed by an even greater number (n+1) of *slightly* lesser harms.** Thus, it seems that one human death can be outweighed morally by 100

cases of near-death (i.e., the condition that is nearly as bad as death).

(2) **<u>Continuity thesis</u>: All types of harms form a continuous spectrum, such that starting from death, there is always a slightly lesser harm.**

(3) **<u>Transitivity thesis</u>: If A morally outweighs B, and B morally outweighs C, and C morally outweighs D, …, and Y morally outweigh Z, then A morally outweighs Z.**

If all theses are true, then the Repugnant Consequence would follow. (And the Inevitability Argument would be true.)

I shall argue that the Repugnant Consequence does not follow. My stance is this: The Continuity Thesis is self-evident. While the Tradability Thesis is not self-evident, it is nevertheless highly plausible. In arguing for the Tradability Thesis, I show that approximation is involved in each step of the argument. But approximations – or judgments based on approximations – are not transitive. Hence, I shall conclude that the Transitivity Thesis is not true without qualifications.

Briefly, my view is that the Transitivity Thesis is true with respect to one-variable comparisons of harm, but not necessarily true with respect to genuine, or irreducible, two-variable comparisons.[8] Consequently, the Repugnant Conclusion does not follow, because two-variable comparisons of harms are involved. (I will say more about irreducible two-variable comparisons below. See Section V.)

II. Why the Tradability Thesis is plausible: The Two-variable Approach & the Approximation Approach

Since I take the Continuity Thesis to be self-evident, I shall only try to show that the Tradability Thesis is highly plausible. Now let me put forward this claim:

> (P) The degree of **<u>m</u>oral** urgency (M) that we should accord to a group of people who need our help is *on a first approximation* a function of two variables or dimensions, namely, the degree of harm (h) to be prevented, and the number (n) of people who would suffer such harm if not saved from it. We have: **M**=**M**(h, n).

[8] A genuine, or irreducible, two-variable problem is one which cannot be reduced into, or become, a one-variable problem.

As I have argued elsewhere,[9] certain type of comparison can be usefully thought of as a problem in two variables. I believe that this two-variable approach might shed light on the present problem, which is also a two-variable problem.

A two-variable problem is simply a problem that consists of two variables or dimensions. The sort of two-variable problems that I discuss below have the following characteristics: First, they are genuine, or irreducible, two-variable problems in that they are not reducible into one-variable problems. Thus, the speed of a sprinter may be a function of her mass and her explosive power. Yet, given a certain mass and a certain explosive power, a sprinter will have a certain speed. Because the variables of mass and explosive power are reducible into the variable of speed, this is not what I call a genuine, or irreducible, two-variable problem.

Let me illustrate with examples what are irreducible two-variable problems. One example occurs in the problem of animal rights. In the case of animal rights, as I have argued elsewhere but can here only state, we are faced with an irreducible two-variable problem, which cannot be reduced into a commensurable scale of utilities.[10] The two variables here are those of (a) type of claim, and (b) intrinsic value or moral status. And when we try to compare (1) our claim to be saved from

---

[9] I argue that the problem of animal rights is a two-variable problem, in "Animal Research, Non-vegetarianism, and the Moral Status of Animals – Understanding the Impasse of the Animal Rights Problem," *Journal of Medicine and Philosophy*, 2002, Vol. 27, No. 5, pp. 589-615, as well as "Toward Quasi-vegetarianism," in Hon-Lam Li and Anthony K. W. Yeung, eds., New Essays in Applied Ethics: Animal Rights, Personhood, and the Ethics of Killing, UK: Palgrave Macmillan, *forthcoming*. I also argue that the problem of abortion is also a two-variable problem in, "Abortion and degrees of personhood: Understanding the Abortion Problem (and the Animal Rights Problem) are Irresolvable," *Public Affairs Quarterly*, 1997, Vol. 11, No.1, pp. 1-19.

[10] See Hon-Lam Li, "Animal Research, Non-vegetarianism, and the Moral Status of Animals – Understanding the Impasse of the Animal Rights Problem," *Journal of Medicine and Philosophy*, 2002, Vol. 27, No. 5, Section VII, pp. 600-605.

an injury, and (2) a dog's claim to be saved from death, we do not know how to make such a comparison – that is, we do not know how to compare whether a human's injury is more serious than a dog's death or not. But we know how to make comparisons of claim in the following obvious cases: First, if one of the two variables is held constant – if, for example, we can save either a person from death, and another person from injury, but not both – then we should save the person from a greater harm, and in this case, the former from death. This intuition can be plausibly explained as follows: Since the variable of intrinsic value is being <u>held constant</u>, the variable of type of harm determines the outcome of the issue. Another example where there is an obvious solution is this: suppose we can save a dog from death or a human from death. We should save the human from death. Again, the variable of type of claim being held constant, it is the variable of intrinsic value or moral status that dictates who we should save.

There is a situation in which a solution to a two-variable problem is obvious, is when <u>both variables are in favor of one party</u>. Suppose we can save either a human being from death, or a dog from injury, but not both of them. Clearly, we should save the human being from death, because both the variable of moral status or intrinsic value, and the variable of type of claim, are in favor of the human being.

This approach of conceiving of the problem of animal rights in terms of two variables might shed light on some of the cases in which we have to decide which group of people to save, because the structures of these problems are identical. Thus, when we have to decide between saving either one person, or two other persons, but not everyone, then the two-variable approach produces the intuitively correct result.

In this example, the variables are those of harm (h) and number (n) respectively. Since the variable of harm is held constant, the variable of number makes the difference. However, if we can save either one person from death, or one thousand persons from blindness, then the two-variable approach will not be able to come up with any solution. For whereas the variable of harm is in favor of the former, the variable of number is in favor of the latter. Hence, we are stuck.

To explain the character of a two-variable problem further, consider the following analogy. We can, in mathematics, solve an equation with one variable (e.g., 2x=4) -- but not equations with two variables (e.g., 2x+y=4), because there is one unknown too many. The moral analogue in the present case is this. Suppose Group A has a certain number of people, each suffering from a certain disease. Group B also has a number of people, and each of them is also suffering from a certain disease. We can solve the problem of competing claims of the two Groups if the diseases are equally severe, but one group has more people. We can also solve the problem if the number of people is the same in both group, but one disease is more severe. In both cases, one variable is held constant and is effect eliminated. *A fortiori*, we can solve the problem if both variables are in favor of one group. However, we do not know how to solve the problem if Group A has more people, but the disease Group A suffers is less severe than that suffered by Group B. We do not know how to deal with the problem, because there is one unknown too many. Note that in both examples, there are too many solutions: In the case of the equation "2x+y=4" there are three solutions if we restrict x and y to be natural numbers,[11] whereas in the present case (or in the case of animal problem), the solution is also under-determined. In both cases, if we look for a unique solution, we are disappointed, since there are too many solutions. A unique solution would exist if we eliminating one of the two variables, and in this sense both cases are undecidable or indeterminate.

This analogy is, however, heuristic and not exact, since all three solutions in "2x+y=4" are good solutions and there is no mathematical reason to look for a unique solution. The following function F would provide a more exact analogy: (1) F is a function in two variables x and y; (2) F is undefined over a certain range R; (3) F is defined over R if x or y is held constant.

Yet, a two-variable problem can be solved, not only in cases where (1) either one variable is being held constant, or (2) both variables go in favor of one party, but also in certain cases **<u>(3) where the method of approximation can be used to arrive at a solution</u>.** By means of approximation, we can sometimes reduce a two-variable problem into a one-variable one, thereby enabling the problem to be solved. The method of approximation, which mathematicians and scientists use, can

[11] The solutions are (0, 4), (1, 2), and (2,0).

also be used profitably by ethicists, if so doing will enable some problems to be solved. To illustrate, consider whether to save a chimpanzee from injury, or a gorilla from death. We have a formally two-variable problem that we cannot resolve, because the chimp has a slightly higher degree of intrinsic value (because of a slightly higher capacity for experience), whereas the gorilla's type of claim is weightier. However, because the chimp's intrinsic value is only marginally higher than the gorilla's, whereas there is a big difference between injury and death, we should save the gorilla from death. Another example: Suppose we can save a human from injury or an oyster or shrimp from death. Who should we save? While we have a formally two-variable problem, the moral status or intrinsic value of an oyster or shrimp must be close to nothing. Hence, by approximation, we can deem the intrinsic value of an oyster or shrimp to be zero. Consequently, we should save the human from injury because her injury is morally weightier than the death of an oyster or shrimp.

The relevance of the method of approximation to the two-variable comparison of harms is this: Surely, we cannot use the method of approximation to

---

[13] When we have achieved this, we might have reasons to believe that we should save a finite number of people from blindness instead of one person from death, if the number is large enough. As Bernard Williams points out, while it is illegitimate to slide all the way down a slippery slope, it is all right to slide a little bit. Has the slide from death to blindness exceeded the limit? Probably not, though I am not entirely certain. For it depends on how slippery the slope is, so to speak. If the slope is very slippery, it would be illegitimate to slide even a little bit. However, the way to decide whether a slope is too slippery or not – or whether the slide from death to blindness has exceeded the limit or not – is ultimately a matter a intuition.

reduce the comparison of one human life and n cases of blindness, because the gap between death and blindness is too wide, for blindness is ***not*** almost as bad as death. But we might wish to consider whether reduction can be achieved by means of ***a series of approximations***. So between death and blindness, there are a finite number of types of harm which, while worse than blindness, are less bad than death. Consider Mary's condition: Mary, who -- unless we save her now -- will become blind, deaf, lose all his limbs, and have much of her mental capacity reduced. Mary's condition will be almost as bad as death. Call her condition one of **near-death.** Suppose we can save either someone from death, or 100 people from **near-death**, but not both. Who should we save? If we use the method of approximation, we can arrive at the result that we should save these 100 people in the following way. Firstly, given that Mary's condition is nearly as bad as death, we can by the method of approximation deem it to be as bad as death. Secondly, since Mary's condition is as bad as death, the two-variable problem is hence reduced into a one-variable problem. Thirdly, the variable of harm being held constant, it is the variable of number that determines the issue. Hence, we should save the 100 people suffering from **near-death**.

Having achieved this result, one is tempted to repeat this process of approximation, by considering whether we should save someone from **near-death**, or

100 people who would otherwise suffer from **near-near death**, which is nearly as bad as **near-death**. Because **near-near-death** is nearly as bad as **near-death**, then we can again use the method of approximation to yield the result that we should save 100 people from **near-near-death** rather than one person from **near-death**. This process of approximation can **perhaps** be repeated until we reach the case where someone will become blind unless we save him.[13]

Now, if I am correct, there exist a finite number of similar claims such that they can outweigh a slightly weightier claim. **Hence, the Tradability Thesis.** But this ***seemingly*** leads to the repugnant conclusion. **I shall argue below that this does not lead to the repugnant conclusion.**

III. Why the Repugnant Conclusion does not follow

In section II, I put forward (P). A variation of (P) is:

(Q) The degree of **moral** urgency (M) that we should accord to a group of people is a function of two variables or dimensions, namely, the degree of ***how <u>u</u>nbearable*** (u) certain harm is, and the number (n) of people who would suffer such harm if not saved from it. So, I propose: M=M(u, n).

Suppose a tyrant imposes on Smith a huge amount, say, 100 units of tax. One might say that it is worse for 2 persons each to have 90 units, and that it is still worse for 4 persons each to have 80 units of tax, and so on. Fairly soon, we will arrive at the conclusion that it is still worse for 4,096 people each to have 1.25 units of tax. Now it seems that this conclusion is false, and if so, why? The reason has to do with (1) how the burden of tax is distributed, and consequently (2) whether one can bear or afford the burden. It seems that it is worse for one person to suffer an unbearable burden than for 4,0960 people each to bear an affordable burden. Now it seems what the total amount of tax the 4,096 people have to pay is unimportant, as long as each of them can afford it.[14] If this is not obvious, we can still thin out the burden such that 32,000 people will each share 0.15 unit of tax.[15]

This conclusion also seems to accord with the account of approximation proposed. If 0.15 unit of tax is trivial or negligible, it can be deemed to be approximately nothing. If so, then 32,000 people each of whom bears approximately no burden is not a morally worse consequence, but in fact a morally better consequence, than one person having to suffer an unbearable burden. The same goes

[14] An analogy would hold between this case and the case of death vs. headaches.)

[15] An analogy with the headache situation is that it is worse for one person to die than for many many people each to have a trivial itch lasting for one second.)

for death and headaches. A headache is trivial or negligible when compared with death. We can deem the headache to be nothing. A huge number of people, each having a trivial harm, is not a worse consequence than one person who is dying. It is self-evident that a minor itch lasting one second, even if had by an astronomical number of people, does not outweigh a person's death. For such a one-second itch is literally nothing. An astronomical number of people, each suffering nothing, are not a worse but actually better consequence than one person dying. These examples show, *via* approximation, that the Aggregability Axiom is false, that a huge number of trivial utilities do not aggregate to outweigh a substantial loss, such as the loss of a life.

However, by the method of approximation, Mary's **near-death** condition is almost as bad as death. Hence, the two-variable problem is reduced into a one-variable problem. Hence, 100 cases of near death can be deemed to be worse than one case of death.

Now imagine that you can save either someone's life, or 10,000 people each having to lose a hand, but not both. Who should you save? We have a two-variable problem that cannot be reduced into a one-variable problem. For one thing, such a reduction would involve too many successive approximations -- and hence seems to be unsolvable. For another, the answer is indeed indeterminate.

So if I am correct, the account based on approximation and the idea of two variables explains (1) why 100 cases of near-death morally outweigh one human life, (2) why a huge number of one-second itches do not morally outweigh one human life, and (3) why it is unclear whether there exists a number of human hands such that they would morally outweigh one human life.

IV. A Thought Experiment

The intuition that many many minor headaches are not as bad as, and do not outweigh, a human being's life can be confirmed in the following thought experiment. Suppose that retributivism is correct, that is, a person deserves punishment to and only to the extent of his culpability in committing a criminal act. Now on any view, one's culpability is determined by (1) one's intention (e.g., whether it is pre-meditated, reckless, foreseeable, or totally unforeseeable and hence unintended), and (2) the gravity of one's act. The gravity of one's act is just the consequence of the act. Now if John, Jack, and Tom all pre-mediate and willfully commit their respective criminal acts, and none of them have any legal (or moral) defenses in their favor. If John causes one person to die – assuming that retributivism is correct – John deserves to die. Now if Jack causes 10,000 persons to suffer near-near-death, it would seem that he also deserves to die. However, if Tom causes a minor headache to each of many people, it would seem that he does not deserve to die. It would be even more obvious that he does not deserve to die if he causes a one-second itch to each of a billion people. If these intuitions are firm, as I think they are, then we have further grounds for believing that many minor headaches do not outweigh a human being's life.

V. Do we have the Sorites Paradox?[17]

Let us return to the observations made at the end of Section III: (1) 100 cases of near-death morally outweigh one human life, (2) a huge number of one-second itches do not morally outweigh one human life, and (3) it is unclear whether there exists a number of human hands such that they would morally outweigh one human life. What do these observations suggest?

The moral phenomenology of these observations suggests that we might be dealing with a Sorites Paradox. A Sorites Paradox has five characteristics: (1) the values or properties of the two poles of a spectrum through which a Sorites Paradox traverses are opposite or reversed (e.g, there is a heap in one end, but not in the other end; someone is bald and not bald); (2) a Sorites Paradox involves many steps; (3) in each step, approximation is involved; (4) somewhere between the two poles there

---

[17] Larry Temkin has argued that the present problem is ***not*** a Sorites Paradox. See his "A Continuum Argument for Intransitivity," Philosophy and Public Affairs 25, no. 3, Summer, 1996, pp. 175-210.

exists a range with vague or indeterminate boundary, and the value or property over this range is itself unclear or indeterminate; (5) there is something vague or indetermindate about the central concept or feature which the Sorites Paradox exploits throughout the steps.

First, the two end-points of a spectrum through which a Sorites Paradox works are self-evidently of opposite values or properties. For example, one grain of sand does not form a heap, whereas one million grains form one. A man with 3,000 hair on his head is not bald; a man with one hair is bald. Similarly, in an empirical type of Sorites Paradox, if one person does not vote, it would not make any difference to the outcome of an election, since one vote would not make any difference; but since this argument applies to everyone who is eligible to vote, it is clearly false that everyone's not voting would not make any difference to the outcome of the election. An analogy in the present case is that 1,000 near-deaths would outweigh 1 death, whereas one billion one-second itches would not do so.

Second, a Sorites Paradox involves many steps. The self-evidently false conclusion to be drawn from these steps (for example: "A man with one hair is not bald") cannot be drawn in a few steps. The same is true of the voting paradox, which consists of as many steps as there are eligible voters. In a similar way, the apparently false conclusion that many minor headaches can outweigh a human being's life has to go through many steps, from death, to near-death, to near-near-death, … to arms, hands, and fingers, to sever headaches, moderate headaches, and minor headaches.

A Sorites Paradox has many steps because of a third characteristic: In each step of the argument in a Sorites Paradox, approximation is involved. To use an example in mathematics: $1.001 \cong 1.002$; $1.002 \cong 1.003$. Only through 1,000 steps will we get to self-evidently false conclusion that $1 \cong 2$. In each step, an approximation is made. Similarly, one hair does not make anything difference as to whether one is bald or not. This is a conceptual issue. The voting paradox says that your vote does not make any difference to the outcome of the election. All three claims here are true, but only approximately so. 1.001 is approximately equal to 1.002, because 0.001 is deemed to be irrelevant. A hair makes a difference to one's baldness, but in only such a small degree as to be imperceptible. A vote makes a difference to the outcome of an election, but only with an extremely small probability. (There were cases in Hong Kong's district elections in which the outcome was separated by only one vote.) The same is true of the comparison between death and near-death; approximation is needed to make the argument go through in a two-variable comparison.

Fourthly, there does not exist a clear boundary between baldness and non-baldness, or between the concept of a heap and that of a non-heap. The non-heap ***gradually*** becomes a heap in a way that is vague or a matter of degree. Similarly, there is no clear boundary between a case where one is winning an election, and a case where one is losing – at least not *a priori*. The same, again, is true of the problem we are considering. While 10,000 cases of near-near-deaths would clearly outweigh a human's life, and clearly no number of one-second itch would outweigh a human's life, there exist a range of harms, such that it is not clear whether there exist a finite number of these harms that are aggregable to outweigh one person's life. Thus, it is not clear on which side hands and arms would fall, because they are probably in the indeterminate range.

Finally, there is something vague or indeterminate in the central concept or feature which the Sorites Paradox argument exploits throughout the steps. This characteristic explains the fourth one, as well as the other characteristics. Thus, the concept of "approximately equal to" is vague, which explains the lack of clear boundary where the relation "approximately equal to" begins to fail to hold. It also explains the change of value from a true relation ($1 \cong 1.001$) to a false one ($1 \cong 2$). So are the concepts of a heap and baldness; this explains the lack of clear boundary between a heap and a non-heap, and between baldness and non-baldness. Moreover, there is vaguenss in the ***concept*** of a heap – or some may say in the heap itself – because this feature explains the other two features. For the vagueness of the concept of heap explains why there are indeterminate cases as to which something is a heap or not. Moreover, by constantly adding a grain of sand into a grain of sand (which is clearly a non-heap) through the range of indeterminate cases, we can see how a non-heap becomes a heap, and hence that two endpoints of the process have contrary properties. Similarly, whether someone will win the election is indeterminate *a priori*. Similarly, there is a good deal of undecidability or indeterminateness in two-variable comparisons, as argued in Section II.

To elaborate on the vagueness or indeterminacy of comparing harms, we must distinguish the following two types of comparison. The first type is a simple one-variable comparison. Death is worse than, and hence morally outweighs, blindness. Blindness outweighs the loss of a foot. The loss of a foot outweighs a headache. There is no vagueness or indeterminacy here, because transitivity holds in such one-variable comparison of harms. (This is a one-variable comparison because only one person is involved – and the variable of n is held constant or eliminated.)

There is another type of comparison, that is, pair-wise comparison of genuine two-variable cases. As I have tried to show earlier in Section II of this paper, indeterminacy sets in when we are comparing a two-variable harm with another two-variable harm: e.g., a person's injury vs. a dog's death, or one case of death vs. 10,000 cases of losing one hand. These are genuine or irreducible two-variable comparisons. Indeterminacy or undecidability sets in. Just as the function F (or alternatively and more roughly the equation "2x+y=4") is under-determined, so are these two-variable comparisons.

Thus, the indeterminacy intrinsic in irreducible two-variable comparisons explains why we have a Sorites Paradox here. For such indeterminacy – rather than vagueness – explains the other features of a Sorites Paradox, and the moral phenomenology, such as why (1) neighboring cases can be compared,[18] (2) indeterminacy sets in when two cases under consideration are farther apart,[19] (3) the two endpoints have opposite properties.[20]

In this regard, the tyrant example clearly displays a Sorites Pardox: First, it is worse to have two persons suffering from unbearable taxation (90 units of tax) than for one to bear it (100 units). Second, it is unclear as to whether it is better or worse for one person to suffer unbearable taxation than for 32 people each to suffer relatively heavy taxation (50 units), because there is indeterminacy in (irreducible) two-variable comparisons. Third, it is, however, clearly better for 4,096 people each to bear an affordable tax (1.25 units) than for one person to suffer an unbearable tax (100 units).

---

[18] For example: 100 near-deaths outweigh one death. A person suffering from near-death outweigh the death of a dog.

[19] Thus, it is unclear whether many hands can outweigh a human life. Also unclear is whether a human's injury outweighs a dog's death.

[20] One example is that an astronomical number of people suffering a one-second itch does not outweigh a person's life, but on the contrary is outweigh by it. Another example: A person's bruise does not outweigh a dog's life, but is outweighed by it.

V. Why two-variable comparisons might not be transitive

To recap: I have tried to block one forceful argument for the view that the Repugnant Conclusion can be arrived at even if the Aggregability Axiom is not assumed. My strategy has been to show that approximation is involved in each step of the two-comparison of harms. Another argument appeals to the possibility that we are dealing with a Sorites Paradox. But both arguments are closely related. **For both approaches forbid the repeated use of approximation in an argument. I take it that this is clear in the two-variable approach. In this approach, approximation is allowed but only up to a point. Approximation cannot be used repeatedly.**

Moreover, the Sorites Paradox also has to do with making approximations. If we gradually pick the hair of someone (who is not bald), and declare that he is not bald, we are making only an approximately true statement. We are making a judgment based on an approximation, namely, that one hair does not matter. But approximate judgments cannot be made repeatedly. In other words, successive judgments that are based on a series of approximation cannot be made legitimately.

Because approximation is not transitive (or judgments based on approximation cannot be repeatedly made), the Transitivity Thesis is undermined at least in the case of irreducible two-variable comparisons.

Let me explain why two-variable comparisons may be different from one-variable comparisons. If A is a faster runner than B, and if B is a faster runner than C, then A is a faster runner than C. This is a one-variable comparison, and the Transitivity Thesis holds here. But consider the following multi-variable comparison. Assuming that A is a better player than B, if and only if A beats B more often than B beats A, we have the following consequences: Borg is a better tennis player than Connors (because Borg regularly beats Connors). Connors is a better tennis player than McEnroe (because Connors regularly beats McEnroe). But McEnroe is a better tennis player than Borg (because he regularly beats Borg). Tennis is a multi-variable game that involves more than one variable: it involves attack, defense, and the ability not to make too many unforced errors. Further, pair-wise comparison is necessary. Therefore, it is possible that A beats B, B beats C, but C beats A. So, my **conjecture** is this:

> **(1) If a comparison is one that involves two or more irreducible variables, and (2) if the comparison must be made pair-wise, then transitivity does not necessarily hold.**

For transitivity to hold in the tennis example, one can relax requirement (2), such that whoever earn most points from grand prix and grand slam events will be the best player in the world. But this would have the strange consequence that the best player winner may regularly lose to someone who is not the best player.

Another thing to note is that a one-variable comparison does not seem to require pair-wise comparison. Whoever is the fastest runner can be determined by placing all eight runners on the same track. There is no need to make pair-wise comparison, because if A runs faster than B, and B runs faster than C, then A runs faster than C. But the same is not true of games which require irreducible variables or dimensions, such as tennis, badminton, basketball, soccer or boxing.

So intransitivity in comparisons seems to require both conditions (1) and (2) be fulfilled. Now it may be asked why the two-variable comparison of harms must be made pair-wise. Why can't they be compared in the same way that the runners be compared – that is, all on one scale? To answer this question, I need only to point out that an insistence that the two-variable comparison of harm be on one single scale is begging the very question at issue – namely, it assumes that all types of harm (in multi-variables) are commensurable on one scale – without argument. I have argued elsewhere[21] that all claims by all types of (human and non-human) animals cannot be put on one scale. Here, I have tried to show that two-variable comparisons of harms and numbers cannot be put on one scale, either. Utilitarians – though not only utilitarians – would want to assume that all harms and all claims can be measured on one scale. But, as I have tried to argue, this assumption would be begging the very question at issue.

VI. Objections and Replies:

The first objection goes like this:

> It is granted that the judgment that 100 cases of near-death outweigh one death is an approximate judgment, and hence that the relation "is approximately equal to" is not transitive. However, if we think that one death is approximately equal to 100 cases of near-death, then we can increase the safety margin by saying that 1,000 cases of near-death clearly outweigh one death, and 1,000 cases of near-near-death clearly outweigh one near-death, etc. Now, given the safety margin, why can't we say that there exists a number, n, such as n cases of headaches (clearly) outweigh one

[21] Hon-Lam Li, "Animal Research, Non-vegetarianism, and the Moral Status of Animals – Understanding the Impasse of the Animal Rights Problem," *Journal of Medicine and Philosophy*, 2002, Vol. 27, No. 5, Section VII, pp. 600-605.

death?

To meet this challenge, we have to understand the problem correctly. The claim I have made is not the epistemological claim that we do not know whether it is justified to say that there exists a number, n, such that n cases of headaches outweigh a death. Nor is it the epistemological claim that we do not know what the number n is. It is rather a metaphysical claim that the number, n, such that n cases of headaches morally outweigh a death simply does not exist.

One may wish to ask, why not? The answer has to do with the question whether the two variables of number and harm are commensurable or not. If they are commensurable, then there must exist a number, n, such that n cases of headaches morally outweigh a death. For that would amount to accepting that utilities dispersed among different individuals, no matter how trivial these utilities are, must be aggregable to outweigh a person's life. But whether the two variables are commensurable is precisely the question. I believe it is possible to resist the view that they are commensurable.

To illustrate the point that two variables are not commensurable, consider this example: Agent-neutral and agent-relative values seem to be incommensurable. Suppose your mother (who is not a swimmer) falls into the water. You can save either her, or a number of strangers (not being swimmers) who also fall into the water, but not all of them. Who should you save? The following are clear: (1) If you are faced with either saving your mother, or another stranger, but not both, not only would you be permitted to save your mother, but you should do so. It would be puzzling if you choose to save the stranger, or even toss a coin. A plausible explanation is that whereas your mother has both agent-neutral and agent-relative values, the stranger only has agent-neutral value. (2) Even if there are five strangers, it is reasonable and surely permissible for you to save your mother. This conclusion should remain unchanged, even if the number of strangers is increased to 50. But indeterminacy begins to set in when the number climb to 5,000, or 50,000, or bigger. Is there a number, n, of strangers such that you would be required to save them rather than your mother. This is a grey area. A plausible explanation is that, contrary to the utilitarian view that agent-relative value is actually a type of agent-neutral value, agent-relative and agent-neutral values are irreducible types of value. We do not know how to compare them, except in clear cases at the two endpoints of a whole range of cases. In this sense, these two values are incommensurable. (3) Despite the fact that they are incommensurable, if you can save either your mother from bruising her arm, or everyone else on earth from death, it would be madness not to save the latter. Again, approximation seems to be operative here. Even if your mother has huge agent-relative value for you, a bruise is nothing, and hence it is nothing even if it happens to be your mother who suffers it. Thus, clear cases exist.

Now, the point that I tried to defend is that the variables of number and harm are incommensurable, while pointing out that in certain cases we can arrive at certain conclusions by comparing states of affairs *via* approximation, such as the conclusion that 100 cases of near-death are morally worse than one death. At least, utilitarians (or those who believe that utilities are interpersonally aggregable) would beg the question against non-utilitarians who think otherwise.

Another objection runs as follows:

> If A morally outweighs B, and B morally outweigh C, … , Z morally outweighs Y, but A morally outweighs Z, then there is no best or worst outcome. And this is very counter-intuitive. At least, you would want to say that an astronomical number of one-second itches is the most tolerable outcome. But you can't even say this, given circular relation of "morally outweighing" from A to Z.

I confess that I don't quite know how to respond. But there seem to be two possible replies. The first one and the simplest one is to admit that there is no best or worst outcome, if all outcomes are considered globally. For only local comparisons can be made. One might find this answer highly implausible, but it seems to me that this is no more implausible than saying that there might no best (or worst) tennis player in the world.

Another approach is look at the problem as follows: We can divide all types of (physical) harms into three regions: (1) serious harms – ranging from death to blindness, (2) moderate harms – from loss of a limb to loss of a finger or toe, and (3) minor harms – from a severe but transitory headache to a one-second itch.[22] We can treat the region of moderate harm as a grey area, since it is unclear whether the loss of a number of hands can outweigh the loss of a life, and unclear whether a number of severe headaches can outweigh the loss of a hand. It is also a region with a vague boundary on each side. My suggestion is this: **No number of minor harms can morally outweigh any serious harm.**

VII. Norcross's argument

I should like to end this paper with a (somewhat speculative) response to Alastair Norcross's argument that many headaches can outweigh a human death.

[22] I am inspired here by T. M. Scanlon, *What We Owe to Each Other*, pp. 238-9.

Norcross's argument goes as follows:-

(1) If one has a headache, and there is no painkiller at home, it would be permissible for one to drive to get some from a drug store.

(2) If it is permissible for Adam to drive to get some painkiller, and similarly permissible for Bernie, Claire, David, Elizabeth, …., to do so, then inevitably someone will get killed in a traffic accident, if a huge number of people independently drive to a drug store to get painkiller.

(3) (Conclusion) It is permissible to trade a human life for a huge number of people's headaches.

The structure of this argument is similar to Nozick's Wilt Chamberlain argument to the extent that we are asked to consider the consequence of a single act and that of a conglomeration of many such acts:

(1') It is permissible for Chamberlain to entertain his fans each for a small fee one evening after work.

(2') If it is permissible for Chamberlain to entertain his fans each for a small fee one evening after work, then it must also be permissible for Chamberlain to do so every evening.

(Conclusion) It must be permissible for Chamberlain to do so even if there is huge income inequality between Chamberlain and his fans.

Thomas Nagel has effectively criticized Nozick's argument. Nagel points out that there is no reason to believe that if a single act is permissible, than the conglomeration of all such similar acts must be permissible, anymore than the reverse is true.

So if Nagel's criticism of Nozick is plausible, one might think that Norcross's argument is suspicious, because both arguments share a similar structure. In fact, I believe that Norcross's argument is even more vulnerable than Nozick's, because Norcross's claim is much stronger. For he wants to argue to the conclusion that (1) not only is it permissible for many people to drive to get painkiller, even if that would mean someone's death in a traffic accident, but also that (2) it is a better consequence for many people to have their headaches alleviated than for one person to die. Now an argument analogous to Nozick's would be to draw an analogous conclusion of (1) – not (2). Even if we accept (1), there is no reason for us to accept (2).

To see this, consider the following counter-example. Assume that (a) it is permissible for each person to devote every weekend to the expensive and dangerous hobby of car-racing – suppose 10% of the people who practice this hobby die annually. Is it permissible for everyone to practice such this hobby? We do not need to answer this question, because Norcross would want to draw a much stronger conclusion, namely, that it is a better consequence for many people to race cars on weekend, even if 10% of those practice the hobby of car-racing die from traffic accidents annually. For those who practice this hobby for more than ten years, 65% die from accidents and only 35% survive at the end of a decade. Is it a better consequence for everyone to practice their hobby of car-racing and have 65% of them die at the end of a decade, or it is a better consequence for no one to practice such a hobby and everyone to survive at the end of a decade? This question does not involve permissibility, since permissibility is not the issue. The issue concerns whether we can infer from a premise regarding permissibility for each to do X (whether or not others do so), to conclusion that it is a better *consequence* for everyone to do X than for everyone not to do X. Apart from car-racing, taking dangerous drugs might be another example. A libertarian would say that it is permissible for any adult to take dangerous drug. What obviously does not follow, which is also obviously false, is the conclusion that it is a better consequence for everyone to take dangerous drugs and to have his or her life shortened by 40%, than not.

Word count (text): 7,290