Journal of Applied Philosophy, Vol. 20, No. 2, 2003

Intransitivity and Future Generations: Debunking Parfit's Mere Addition Paradox

KAI M. A. CHAN

ABSTRACT *Duties to future persons contribute critically to many important contemporaneous ethical dilemmas, such as environmental protection, contraception, abortion, and population policy. Yet this area of ethics is mired in paradoxes. It appeared that any principle for dealing with future persons encountered Kavka's* paradox of future individuals, *Parfit's* repugnant conclusion, *or an indefensible* asymmetry. *In 1976, Singer proposed a utilitarian solution that seemed to avoid the above trio of obstacles, but Parfit so successfully demonstrated the unacceptability of this position that Singer abandoned it. Indeed, the apparently intransitivity of Singer's solution contributed to Temkin's argument that the notion of "all things considered better than" may be intransitive. In this paper, I demonstrate that a time-extended view of Singer's solution avoids the intransitivity that allows Parfit's* mere addition paradox *to lead to the repugnant conclusion.*

However, the heart of the mere addition paradox remains: the time-extended view still yields intransitive judgments. I discuss a general theory for dealing with intransitivity (Transitivity by Transformation) *that was inspired by Temkin's sports analogy, and demonstrate that such a solution is more palatable than Temkin suspected. When a pair-wise comparison also requires consideration of a small number of relevant external alternatives, we can avoid intransitivity and the mere addition paradox without infinite comparisons or rejecting the* Independence of Irrelevant Alternatives Principle.

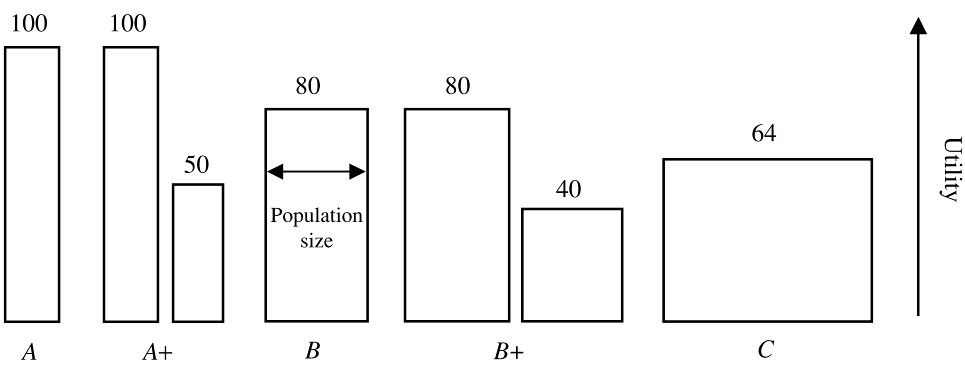
Disagreement regarding the nature and substance of our duties to future persons is widespread. Yet never has a consensus on this issue been so important. Global environmental changes are occurring more rapidly than ever with our blooming human population and warming climate. Sole possessors of the decision-making authority, we stumble through this moral minefield with nary an idea of our proper course.

If the general populace hopes to gain enlightenment from the "experts" — ethicists — regarding duties to future generations, they risk disappointment. This area of ethics is mired in paradoxes. It appeared that any principle for dealing with future persons encountered Kavka's *paradox of future individuals* [1] or Parfit's *repugnant conclusion* [2], or else committed an indefensible *asymmetry* [3]. In 1976, Singer proposed a utilitarian solution that seemed to satisfactorily deal with the above trio of obstacles [4], but Parfit so successfully demonstrated the unacceptability of this position [5] that Singer later abandoned it [6]. Indeed, the apparently intransitive preferences of Singer's solution and of common intuition [7] have led Temkin to argue that the very notion of "all things considered better than" may be intransitive [8]. In this paper, I demonstrate that a time-extended view (TEV) of Singer's solution avoids the intransitivity that allows Parfit's mere addition paradox (MAP) to lead to the repugnant conclusion.

However, the heart of MAP remains: even a TEV of Singer's principle sometimes yields intransitive judgments. Building upon the way TEV fails to elude MAP, I will discuss a general theory for dealing with intransitivity that was inspired by Temkin's sports analogy. I will demonstrate that such a solution is more palatable than Temkin suspected: it does not require infinite comparisons or the rejection of the *Independence of Irrelevant Alternatives Principle* and presents a workable solution to MAP and intransitivity.

Mere Addition Paradox (MAP)

Intransitive rankings of states of affairs results in Parfit's MAP, which indirectly leads to the repugnant conclusion. This paradox has been illustrated with an imaginary population of one million people at a utility value of 100 [9] (example 1, situation *A*; shown below). Given the choice, it seems impossible to justify not adding an additional million people at a utility value of 50 (which would result in *A*+), since these people would not exist otherwise and would appreciate existence [10]. But if we allow the additional people in *A*+, most would agree that we ought to remove the disparity between the two groups, if the gain by those worse off exceeded the loss of the better off. Assuming that this is possible, we soon have two million people at utility 80 (*B*). From *B*, we can be similarly persuaded to move to *B*+, then *C*, from *C* to *C*+, then *D*, and so on [11]. We eventually reach the *repugnant* situation at *Z*, where an immense population leads lives barely worth living. In the simple form of the repugnant conclusion, we would immediately prefer *Z* to *A* because of a greater total utility. In MAP, we consider *Z* grievously worse than *A*, but nonetheless reach *Z* in a series of steps. The path is laid by an apparent intransitivity: *A* is better than *B*, but *A*+ is not worse than *A*, and *B* is better than *A*+; if we start at *A*, we'll quickly move through *A*+ to *B* — a worse state.



Different principles fail to navigate the gauntlet of paradoxes in different ways. The "total" form of utilitarianism [12] — which seeks to maximize total utility, regardless of population size — immediately falls victim to the repugnant conclusion. By this principle, *Z* is better than *A* because 33.55 trillion people ($2^{25} \times 1$ million) at an average utility of 0.38 ($0.8^{25} \times 100$) have a much higher total utility (12.75 trillion) than a million people at an average utility of 100 (100 million). The "average" form of utilitarianism — which seeks to maximize average utility, regardless of population size — also fails immediately. It indefensibly prohibits the creation of additional people — even if they are fully distinct and separate — if they lower the global average utility [13].

Singer's proposed solution lies somewhere between the two extreme forms of utilitarianism presented above. He stated it in a variety of different forms as he sought to avoid the paradoxes above. He first suggested, "A population policy is right only if it does not make those who exist or will exist independently of the policy less happy than they would otherwise have been" [14], but recognized that such a principle fails to deal with the paradox of future individuals. "[T]hose who exist or will exist" fails to include people whose particular identity were affected by the decision, such as in Parfit's *handicapped child* example [15]. Since even slight changes in the timing of conception can change identity, the above principle left out many future people. Singer therefore moved to his principle 4': "If a possible future state of affairs is a world of P people whose total happiness is T, it is wrong to bring into existence any greater number of people, P + N, such that no sub-group of P + N contains P people with a total happiness equal to or greater than T." This phrasing tells us what we should not do, but not always what we should do. Reading Singer's 1976 chapter suggests that he accepted the following principle: *given two alternative futures, we ought to prefer the one with a higher total utility for the best-off* P *people (the* core group*), where* P *is the smaller population size.*

Considering only the next generation, Parfit demonstrated that Singer's principle suffers from the same intransitivity that embodies MAP. When comparing alternatives with A, the core group is the best-off 1 million (P), and their total happiness in A is 100 million. A+ has a subgroup of 1 million with an equivalent happiness, and has an additional million with a lesser — but still positive — happiness. We therefore seem to prefer A+ to A. But once we get to A+, we prefer B, since P for the comparison between A and B is 2 million, and the total happiness of the best-off P in B is 160 million rather than 150 million. The intransitivity arises because B has no subgroup of 1 million with a higher total happiness than A (100 million), and is thus deemed inferior to A, although it is better than a situation (A+) that is superior to A.

This susceptibility of Singer's solution to MAP appears to arise from several sources. First, although we care about generations in the more distant future, Singer's principle does not account for them in its judgments. Second, future generations may make decisions that are undesirable from our perspective because their set of alternatives differs from ours (because of Singer's plausible assumption that there is great cost to reducing population size, once we increase the population size from 1 to 2 million, a population of 1 million is no longer an option for the next generation). Third, the core group varies between comparisons: it is the best-off million when comparing A with A+ or B, but the best-off two million when comparing A+ with B.

Recognizing this route to the repugnant conclusion, perhaps we can avoid it by addressing the first two possible sources of paradox: by including the welfare of distant future generations in the preferability of options, and by accounting for the changing set of alternatives.

A Time-Extended View (TEV)

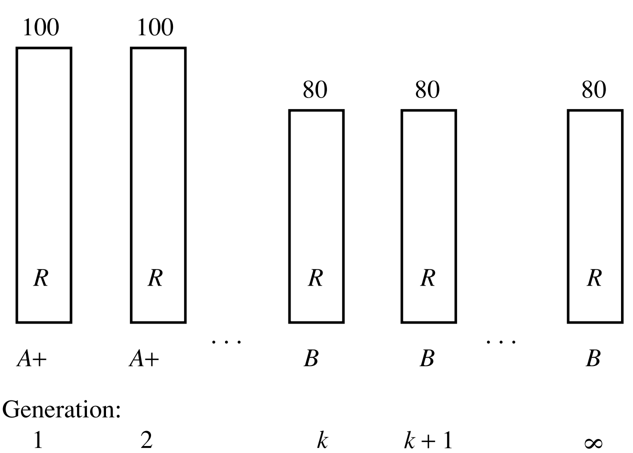
Consider a TEV of Singer's principle. I shall assume that we are concerned with the core group of each subsequent generation, where the core group is the best-off P

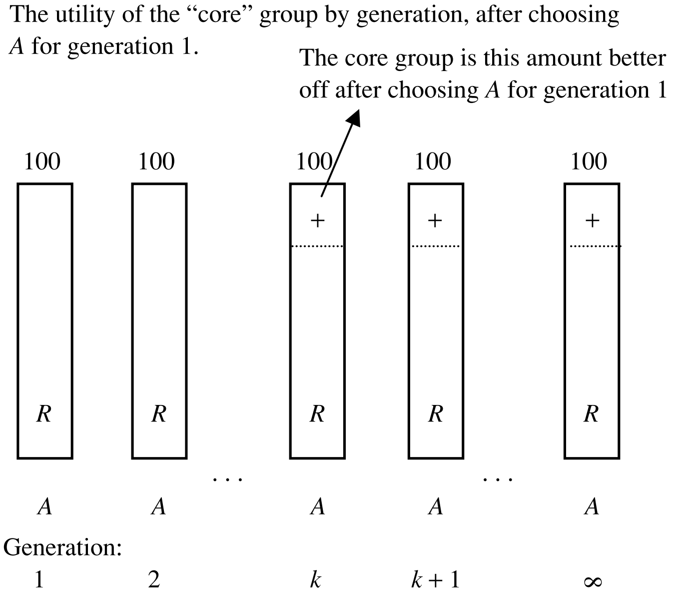
people and P is fixed at the smaller of two population sizes in the first generation [16].

Contrary to Parfit's bleak assessment, a TEV of Singer's principle does not lead to the repugnant conclusion (see diagram below). It is true that the decision to move to $A+$ is not worse for the *first* subsequent generation, but our decision impacts *many* subsequent generations and all generations deserve consideration [17]. The disparity that we create in $A+$ puts the onus on future generations to reduce the unfair discrepancy at the first possibility. This equalization implies B, which reduces the utility of the core group in all subsequent generations (given Singer's and Parfit's assumption that we cannot reduce the population size). We do not actually harm the first core generation in $A+$, but we impose upon them a debt that obliges them to allow inequality (by continuing the asymmetric relation with non-core people) or harm subsequent core generations (by benefiting the non-core people in moving to B). We can plausibly assume that some future generation will remove the disparity to benefit the group they see as their core group, thereby diminishing the welfare of all further future generations of our core group. Because of the immense number of generations adversely affected in this manner by our decision to move to $A+$, we should not make such a move. Contrary to claims of Singer and Parfit, $A+$ is *worse* than A. Thus, in light of all future generations, we shed the earlier intransitivity: situation A is better than B, which is better than $A+$ ($A > B > A+$).

Some readers will feel uneasy with the manner in which the TEV deals with example 1. The logic initially seems similar to Singer's proposed solution to the intransitivity that advocates the move from A to $A+$ to B: prohibit the initial move to $A+$ because one knows that from $A+$ one would prefer B and one does not wish to do so (because one prefers A to B) [18]. Parfit has convincingly argued that such a solution is an unsatisfactory solution to intransitivity, unless one can justify favoring some judgments (A over B and B over $A+$) over others ($A+$ over A) [19]. An even more challenging example (example 2) renders this discomfort more conspicuous, since there are people

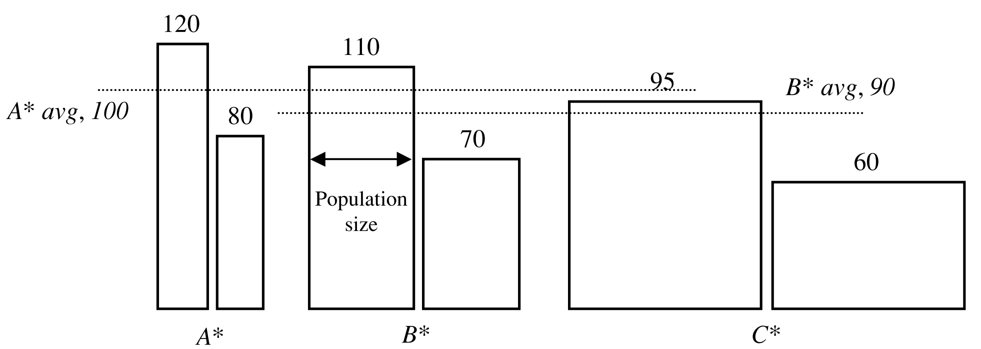
The utility of the "core" group by generation, after choosing $A+$ for generation 1. At generation k, the opportunity arises to remove the disparity, which the "core" group is obliged to do.





suffering in the initial situation, and one feels obliged to help them [20]. Some find this example more difficult because $B^\star$ seems clearly preferable to $A^\star$, whereas $A+$ seems merely "not worse" than A. In example 2 our world starts with an asymmetry that — by some unfortunate twist of fate — we can only eliminate by adding more people, who are themselves comparatively disadvantaged.

Consider the various generations of the "core" group (those who would have existed anyway, interpreted by Singer as the best-off P people, where P is the number of people in $A^\star$). The average *per capita* utility of the best-off P will be 100 if we choose $A^\star$ for the next generation, as opposed to 110 if we choose $B^\star$. The TEV's crucial contribution is that it also considers the "core" group of subsequent generations. But how can we consider those generations if we can't be certain of their fate? The best scenario for "core" groups through time would be for all future generations to choose $B^\star$, thereby securing an average core-group *per capita* utility of 110 for time immemorial. This would seem to condone the move to $B^\star$, but it opens



the door for the next generation to make the same calculation given *their* core group ($2P$ people) and so move to $C^\star$, thereby succumbing to the full mere addition paradox.

Since we cannot force future generations to act as we wish, an appropriate *population policy* would be to act in a way that maximizes the utility of the core groups of future generations, given our knowledge of how each future generation will act. Because the TEV is not concerned merely with single-generation outcomes of the next generation, but with the outcomes of all future generations, we must sometimes prohibit a move to a preferable (from a static perspective) outcome in order to improve the preferability over the long term. Thus, if we knew that future generations would increase population size in order to maximize the utility of *their* children's generation, we would elect not to bring about $B^\star$ in order to slow the world's progress to $C^\star$ and beyond. The trajectory that remains at $A^\star$ is preferable to that which moves to $B^\star$. The table below summarizes the two alternative trajectories.

	Population position and average utility of best-off *P* at *t* =			
Choice from *t* = 0	**1**	**2**	**3**	**4**
Stay at $A^\star$	$A^\star$, 100	$B^\star$, 110	$C^\star$, 95	$D^\star$, 80
Move to $B^\star$	$B^\star$, 110	$C^\star$, 95	$D^\star$, 80	$E^\star$, 65

Although moving to $B^\star$ is better for the core group of generation 1, that benefit is more than offset by a loss to the best-off P of generation 2. Ever after that, the decision to stay at $A^\star$ has benefits for the core groups of those generations. The situation is somewhat more complicated if we are not sure how subsequent generations will act. But so long as there is some probability of future generations acting in the short-term interests of the subsequent generation, it will eventually lead to a net benefit to remain at $A^\star$, although the time at which this strategy pays off will be further in the future the smaller the chance that future generations will increase their population. If we had reason to believe that every tenth generation would act to maximize the short-term gain (similar to a probability of 0.1 for each generation), we would expect the following utilities for the two trajectories (stay at $A^\star$ and move to $B^\star$):

	Population position and average utility of best-off *P* at *t* =			
Choice from *t* = 0	**1–10**	**11–20**	**21–30**	**31–40**
Stay at $A^\star$	$A^\star$, 100	$B^\star$, 110	$C^\star$, 95	$D^\star$, 80
Move to $B^\star$	$B^\star$, 110	$C^\star$, 95	$D^\star$, 80	$E^\star$, 65

In the short term (the first 20 generations), $B^\star$ is the better choice. However, remaining at $A^\star$ eventually pays off. Accordingly, the TEV of Singer's principle avoids the repugnant conclusion: it only condones moving to $B^\star$ when there is no possibility of $C^\star$, $D^\star$, or $Z^\star$.

Although this TEV of Singer's principle avoids the most insidious side effect of MAP (the repugnant conclusion), troublesome aspects of MAP remain. Imagine knowing that future generations will not be able to alter their population size from whatever we choose for the next generation. Comparing the welfare of the best-off P people in each generation, $B^* > A^* > C^*$, $A^* > D^*$, . . . , $A^* > Z^*$. Yet, when we compare B^* directly with C^*, we consider the welfare of the best-off $2P$ people, and prefer C^*. In combination with $B^* > A^* > C^*$ from above, our judgments are intransitive. In such a situation, TEV offers *no guidance* in choosing an alternative.

It seems that addressing the first two factors that appear to cause the repugnant conclusion fails to solve the problem of intransitivity and MAP. Perhaps addressing the third factor, the changing size of the core group, will allow us to avoid MAP.

The Hard Core Principle

Any principle that relies exclusively on pair-wise comparisons that have pair-specific standards of comparison will be vulnerable to the same paralyzing intransitivity as above. In the comparison between A and $A+$ or B, Singer's principle considers the welfare of the best-off million, but when comparing $A+$ and B it considers the welfare of all two million. The best-off million fare better in $A+$ than B, but the best-off two million fare better in B than $A+$ (on average). If we seek to avoid the conflict between judgments, we can keep the standards of comparison fixed, and compare all alternatives on the basis of the same number of people. Parfit discussed this slight revision of Singer's stated theory, where the core group for all comparisons was the minimum population size among all possibilities. He called it "the Hard Core Principle" [21].

According to the Hard Core Principle, we should prefer $A+$ when choosing between A, $A+$, and B. Our ranking would be clear and transitive: $A+ > A > B$. Yet it seems unfair to judge $A+$ better than B on account of the best-off million being slightly better off, given that the worse-off million are considerably happier in B. Furthermore, we would still prefer $A+$ to B even if the worse-off million in $A+$ would have lives barely worth living (say utility 1, where 0 is not worth living) and everyone in B would be nearly as happy as the best-off in $A+$ (utility 99). It is difficult to justify such a conclusion on the grounds that there is another possible future, in which the population size is a million. It is also difficult to justify the underlying reasoning, by which we ignore the welfare of large numbers of people in both of two alternatives because of some feature of a different alternative. Adding to its troubles, the Hard Core Principle also leads to the repugnant conclusion by first condoning $A+$ from A, then $B+$ from $A+$, $C+$ from $B+$, etc.

Thus is the crux of MAP: if we allow our standards of comparison to vary, our judgments may conflict; if we apply a fixed standard of comparison, our judgments seem clearly wrong. Recognizing this important result, some philosophers have suggested that MAP has ramifications beyond population policy. For example, Larry Temkin seized upon our intransitive judgments regarding population additions to suggest that the whole notion of "all things considered better than" (**ATCBT**) might be intransitive [22]. There has even been mention that this intransitivity might undermine the entire enterprise of practical reasoning [23].

Building Upon Temkin's Sports Analogy: A Theory for Transitivity

Fortunately, there is another solution to intransitivity in "better than". In Temkin's seminal contribution on intransitivity, he recognized that there were occasions in which ATCBT was transitive despite intransitivity in a related concept. He provided the illuminating example of sports teams, where it is possible that "the first-place team consistently beats the second-place team, which consistently beats the third-place team, which consistently beats the first team". In one respect — on the basis of the standards of comparison relevant for their head-to-head comparison — the third-place team is better than the first-place team. Nevertheless, we can coherently judge that the first-place team is better — all things considered — than the third-place team, on the basis of it consistently beating more teams.

Moving beyond the analogy is instructive. The key contribution of the analogy is the demonstration that *head-to-head comparisons* may not provide unimpeachable evidence of relations along a *scale of value*. In terms of the sports analogy, "consistently beats" or "is better than" does not imply "is all things considered better than". If the scale of value of interest is in essence comparative, we will not be able to observe this scale of value directly, but must somehow summarize information from head-to-head comparisons. Team sports are perfect examples of such cases, since quality of play has little meaning without competition between teams. Each relation of "consistently beats" is only one component of the "all things considered" quality of the team. Different components of quality (standards of comparison) are relevant in different comparisons, but this does not impede the creation of a single transitive scale of value.

In order to see how the characteristics of different alternatives can alter the standards of comparison of direct comparisons, imagine two basketball teams: X and Y. X has a great offensive center, but Y has a big, strong center who effectively defends other centers. Y's center nullifies X's center, shifting the game's focus to outside jump-shots. Because Y's guards are average jump-shooters and X's are slightly below average, Y consistently beats X. Nevertheless, in play with other teams, X's center often dominates the inside game, largely covering up the team's unremarkable outside shooting and leading his team to many victories. Y wins some and loses some. All things considered, X is the better team.

With respect to varying standards of comparison, sports are an appropriate analogy for the population policy examples in question. When comparing *A* with either *A*+ or *B*, it seems relevant to consider only a million people, because any additional people in *A*+ have little ground to complain, since they wouldn't have existed in *A* (and no one would have existed in their place [24]). Yet, when comparing *A*+ with *B*, we should consider the welfare of all two million people, since all two million people have a parallel in the other world. As with the sports analogy, the relevant standards of comparison vary from comparison to comparison.

We can phrase a theory of transitivity based on this sports analogy. The problem posed is how to derive an appropriate scale of value for a set of alternatives, when the standards of comparison vary. The varying standards of comparison implies that the value is in essence comparative, so it is not possible to simply calculate the value by measuring a characteristic of each alternative (e.g., time for the 100m-dash), or even by calculating a linear transformation [25] of a set of characteristics. If this were possible, direct comparisons between alternatives would not be intransitive the way that they are for population policies and team sports: linear transformations must result in transitive values.

We can use the transitivity of linear transformations to obtain a transitive scale of value from intransitive comparisons by taking a linear transformation of each alternative's performance in a pre-decided set of head-to-head comparisons. I call this theory *Transitivity by Transformation* (**TbT**).

Overcoming Temkin's Objections

TbT enjoys considerable attractiveness, as it is capable of ensuring transitivity of ATCBT, and initially seems appropriate for the population policy dilemma in question. Yet Temkin encountered apparently insurmountable obstacles to any application of his sports analogy, such that he dismissed the analogy as a potential source of a solution to the problem of intransitivity. Since TbT was inspired by the sports analogy, it is worth considering Temkin's obstacles.

The Problem of Infinity

First, Temkin pointed out that in sports, ATCBT refers to how well a team fares against the small, fixed set of alternatives embodied by the other teams in the league. In contrast, when considering alternative future trajectories for a population, there are effectively an infinite number of possibilities. If we attempt to summarize all head-to-head comparisons in order to rank possibilities according to the scale of value representing ATCBT, we run into the "problem of infinity". Since the range of possible population policies is actually a continuum, there are effectively infinitely many policies. Any method of choosing among these possible policies that requires a head-to-head comparison between each possibility and every other possibility clearly encounters difficulty.

This obstacle does not afflict TbT, although it may make it impractical to measure value in some cases. The obstacle only has force when the transformation in question involves many comparisons, which is not always the case. Perhaps in team sports, the concept of ATCBT would require comparison with every possible team, not merely with the existing teams. This would imply that we can never truly measure relations of ATCBT between teams, since we can never assess their performance in competition with every other team. However, for most purposes it is sufficient to measure quality of play by assessing performance of teams in competition with other existing teams. Perhaps this measure of quality does not represent ATCBT, but relevant-things-considered-better-than, RTCBT. Thus, the quality conveyed by "first-place team" does not represent ATCBT and so does not mean that the team is all-things-considered best, but represents this somewhat lesser concept and signifies that the first-place team is best-in-the-current-league-context. Yet, for most purposes we can accept RTCBT as a viable approximation of ATCBT.

In other cases, ATCBT might not require consideration of all pair wise comparisons. In cases like population policy, alternatives may have several components and overall desirability is determined by comparisons based on each component (that is, each population group is a component, and the overall desirability of alternatives is determined by the relative welfare of each group in different alternatives). In such cases, we may obtain a desirability value by comparing the welfare of each group with the best possible welfare of that group in possible alternatives. Clearly, such a method would not require an infinite number of comparisons.

Independence of Irrelevant Alternatives

Temkin's second reason to reject the approach suggested by his sports analogy was that accepting the approach would require us to give up a principle that he holds dear: the *Independence of Irrelevant Alternatives Principle* (**IIAP**) [26]. In Temkin's words, "For any two situations, A and B, to know how A compares to B all things considered, it is, at least in principle, sufficient to compare them directly in terms of each of the factors we care about." Not only did Temkin find this phrasing of IIAP tremendously appealing, he also argued that rejecting it also implied great difficulty in practical decision-making. According to Temkin, IIAP allows us to find the best of n job candidates with $n - 1$ comparisons; without it, we may require $(n \div 2) \times (n - 1)$, which is a great many more.

However, IIAP simply does not apply to cases when the factors that we care about are comparative in essence. With regret, it doesn't only matter how much we have, what matters is how much we have relative to how much we would have had under other circumstances. Consider the case of regret. Here, the relevant characteristics are not limited to how much we have, but also includes comparisons with other possible alternatives, such as *A*+. Put simply, when alternatives are not irrelevant, pair-wise comparisons are also not independent of other alternatives. Thus, the sports analogy and TbT don't require the rejection of IIAP: pair-wise comparisons may require appeal to external alternatives, but only if these alternatives are *relevant*. Because it allows the consideration of external alternatives, TbT coincides well with psychological observations of decision-making [27].

Ethical questions of population policy are in essence comparative: the standards of comparison vary between comparisons. Ethical questions in general often involve the comparison of welfare in one alternative with welfare in other alternatives. When all the same people are involved in the different alternatives, the standard of comparison does not vary: regardless of which alternatives are in question, the welfare of the same people is at stake. Conversely, people who are not present in the other population policy alternative (when there is also no parallel person [28]) plausibly have less justification to claim unfair treatment than those people who would exist (or who would be represented by a parallel person). Because of the varying standards of comparison in population policy questions, we should expect intransitive direct comparisons.

External alternatives are often *not* irrelevant: we often evaluate the relative desirability of two options by appeal to other options. For example, if I am attempting to select a major at university, and am performing a head-to-head comparison between economics and biology, I may consider that my university is the best ranking institution for economics, but that another university has a far superior biology programme. I would consider this fact even if I were presently only attempting to decide between two majors at my present university: if I majored in biology, I would continually regret my choice of university, while if I majored in economics, I would live an entirely different life without a serious interest in biology. Although many philosophers and psychologists consider it irrational to appeal to other alternatives, we frequently make decisions in this manner, for comprehensible reasons [29].

Just as the regret that I would feel in a second-rate biology programme — if a first-rate programme were possible — has a real effect on me, the fact that the additional people in *A*+ would have been happier if they had been in *B* has a real impact on the desirability of *A*+ relative to *A*. If it were not possible to have a more desirable distribution of happiness — as in *B* — the benefits of life to the additional people in *A*+ would be undiminished, and we should prefer *A*+ to *A*. But if *B* is possible, then it is not clear that *A*+ is better than *A*, *all things considered*. In *A*, everyone who exists is grateful for our decision; in *A*+ half of the people have reason to complain about the population policy chosen. To say that the presence of *B* as a feasible alternative should have no effect on the relative desirability of *A* and *A*+ is to fail to recognize that humans are creatures who cannot help but consider what might have been.

Yet when I consider my happiness in economics *vs.* biology, I do not consider my happiness in the biology programme of *all* the other institutions. This is fortunate, since such a procedure would entail the problem of infinity. But it does not seriously affect my decision that I could also have attended a third university, with a biology programme intermediate between the best feasible choice and my present institution, or a fourth university, without a biology programme at all. Rather, I would consider the best available alternative for both biology and economics. Similarly, the additional people in *A*+ would not be strongly affected by the presence of an alternative ~*B* where the best-off million have an average happiness of 85 and the worse-off million have an average of 75. Thus, although some alternatives are relevant, we need not perform head-to-head comparisons between every pair of alternatives. Instead, I suggest that we make a single comparison for each interest of each alternative: compare the satisfaction of each interest (each person's interest) with the maximal satisfaction of that interest (the parallel person's [30] interest) in other alternatives.

It is therefore possible to circumvent Temkin's most serious objections to the sports analogy, and therefore to TbT. I propose the following application of TbT to population policy ethics: preferability of alternatives should be based on comparisons of the welfare of each individual with the best-case scenario for that individual, or for whomever would exist in his/her place. Given such a structure, we can easily avoid the infinity problem by a three-step process: (1) identify the maximum welfare for each group of parallel individuals; (2) for each alternative, calculate the total denied benefits as the sum of the difference between the maximum welfare and the realized welfare; (3) rank alternatives by increasing all-things-considered desirability (the converse of total denied benefits), with the best alternative having the smallest value. This structure is consistent with IIAP and observations of decision-making from psychology.

A full elucidation of the concepts of contingent interests and parallel existence is beyond the scope of this paper [31]. Their use must be fully justified in order for them to be incorporated into an ethical theory. At present, it suffices that a linear transformation of comparisons between the welfare of each individual and that individual's maximum parallel welfare would avoid the problem of intransitivity, without incurring the difficulties of infinite comparisons for each alternative. Contrary to previous claims, this proposed solution to the mere addition paradox does not require the rejection of the *Independence of Irrelevant Alternatives Principle*.

Conclusion

Considering the importance of more distant future generations, it appears that Singer's population principle had more promise than he realized. Although a time-extended view of his principle yields intransitive judgments, it avoids the aspects of Parfit's mere addition paradox (MAP) that lead to the repugnant conclusion. The time-extended view allows us to recognize that the creation of extra beings is not "mere addition" if it has consequences for generations beyond the next. The judgment "all things considered better than" must include these distant consequences, so we avoid the mere addition paradox by giving up one of the three intransitive judgments that comprise MAP: when *A*+ (for the next generation) leads to *B* (for later generations), the preferability of *A*+ to *A* is purely short term. In the long term, *A*+ is better than *A*.

However, two problems stem from Singer's principle's pair-wise comparison of the utility of the best-off *P* individuals, where *P* is the smaller population size. First, the framework of pair-wise comparisons yields sometimes-paralyzing intransitivity, where there is no best option. By building upon Temkin's work on intransitivity, I have phrased a theory for obtaining transitive relations of "all things considered better than" from linear transformations of intransitive relations of "better than in isolation". Using this *Transitivity by Transformation* theory, one can overcome Temkin's objections about the need for infinite comparisons and the necessary rejection of the Independence of Irrelevant Alternatives Principle.

Applying *Transitivity by Transformation* to population policy entails comparing each person's realized welfare with the maximum welfare that s/he — or the person who would have existed in his/her place — would have enjoyed among the available alternatives. The theory's remaining difficulty, of capturing concern for these "parallel" people, requires separate treatment.

The search for a generally successful theory of beneficence — Parfit's "Theory X" — continues, but the ways that Singer's principle fails offers important clues about the construction of an appropriate population principle. Since an appropriate principle must successfully navigate MAP, a promising starting point is the aforementioned application of *Transitivity by Transformation.*

Kai M. A. Chan, Center for Conservation Biology, Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020, USA. kc@kchan.org

Acknowledgments

I would like to thank Peter Singer for his invaluable comments and guidance at every stage of the development of this paper and Simon Levin for his continual support in my ethical endeavours. I would also like to thank Penny Chan, Stephen Gardiner, Brent Howard, Simon Keller, Ljuba Kovacic, Helen Nissenbaum, Joshua Plotkin, Larry Temkin, Martin Wilkinson, and an anonymous reviewer for their invaluable comments on the various incarnations of this paper and the ideas within it. Although this paper has benefited greatly from their suggestions, I accept sole responsibility for the remaining errors that it may contain. Funding for this research was made possible through a Princeton Environmental Institute (PEI) Science, Technology and Environmental Policy (STEP) Fellowship, a Heinz Foundation Environmental Scholarship, a Natural Science and Engineering Research Council (NSERC) Post-Graduate Scholarship and a Julie Payette-NSERC research scholarship; funds for travel to present this talk at the Australasian Association of Philosophy conference were generously provided by the Ford Foundation through the STEP programme.

NOTES

[1] G. S. Kavka (1982) The paradox of future individuals, *Philosophy and Public Affairs*, 11, 93–112.
[2] D. Parfit (1982) Future generations: further problems, Philosophy and Public Affairs 11, 113–172, 142.
[3] Ibid, 148.
[4] P. Singer (1976) A utilitarian population principle, in M. D. Bayles (ed.) Ethics and Population (Cambridge, MA, Schenkman).
[5] Derek Parfit, On doing the best for our children, in M. D. Bayles (ed.) Ethics and Population (Cambridge, MA, Schenkman); see P. Singer (1993) Practical Ethics (Cambridge, UK, Cambridge University Press), pp. 128–131.
[6] Singer, op. cit.
[7] Parfit, Future generations: Further problems.
[8] L. S. Temkin (1987) Intransitivity and the mere addition paradox, *Philosophy and Public Affairs*, 16, 138–187.
[9] I disapprove of the use of numerical 'utilities', as I believe they fail to accurately represent the complexity of our responsibilities to beings. Nevertheless, I adopt them here for the sake of historical convention.
[10] I exclude duties to non-human organisms from these calculations. I set these important considerations aside for the time being in order that we can reach agreement regarding duties to future humans without complications. Establishing compelling duties to non-human organisms represents tomorrow's battle.
[11] Recall that in these examples, taken from Singer's 'A utilitarian population principle' and Parfit's 'On doing the best for our children', actions that reduce population size are prohibited.
[12] Singer, Practical Ethics, 103.
[13] L. W. Sumner ((1978) Classical utilitarianism and the population optimum, in R. I. Sikora and B. Barry (eds.) Obligations to Future Generations (Philadelphia, PA, Temple University Press)) and R. Scott, Jr. ((1978) Environmental ethics and obligations to future generations, in R. I. Sikora and B. Barry (eds.) Obligations to Future Generations (Philadelphia, PA, Temple University Press)) provide several other convincing reasons why the average form of utilitarianism is untenable.
[14] Principle 3 in Singer, A utilitarian population principle, 87–88.
[15] D. Parfit (1976) Rights, interests, and possible people, in S. Gorovitz (ed.) Moral Problems in Medicine (Englewood Cliffs, NJ, Prentice-Hall).
[16] Note that this number P shall apply to all generations. This approach deserves consideration as a possibility: since we are assuming that generations cannot reduce population size, P is the minimum size possible through time. An alternate approach would be allow P to change through time, at each generation being the smaller of the two population sizes existing at that time. A moment's thought shows that this approach would lead to the repugnant conclusion in a series of steps.
[17] David Boonin-Vail attempted to take future generations into account with his 'Population Choice Principle' (see (1996) Don't stop thinking about tomorrow: two paradoxes about duties to future generations *Philosophy & Public Affairs*, 25, 267–307). This principle asks us to weigh alternative choices of population policies in the next generation on the basis of the utilities that would result if resources were distributed in the most equitable manner possible. While I am sympathetic to this approach, the principle requires additional justification not required for a TEV, which merely accounts for the welfare of all people affected by our decisions. Since his principle achieves no more towards a solution of MAP than TEV, I do not see the necessity of such an approach.
[18] Singer, A utilitarian population principle, 96–7.
[19] Parfit, On doing the best for our children, 108.
[20] Parfit, Ibid, 105.

[21] Parfit, Ibid, 106.

[22] Temkin, Intransitivity and the Mere Addition Paradox.

[23] L. S. Temkin (2000) An abortion argument and the threat of intransitivity, in R. Crisp and B. Hooker (eds.) Well-Being and Morality: Essays in Honour of James Griffin (New York, NY, Clarendon Press).

[24] For the importance of this consideration, see Kai M. A. Chan, (submitted) Duties to future persons: solving Parfit's non-identity problem requires a *Plaintiff*-Regarding Principle, *Philosophy & Public Affairs*.

[25] Linear transformations are of the form $V = ax + by + c$, where V is the value in question, x and y are the relevant variables/characteristics, and a, b, and c are constants.

[26] Originally from Kenneth Arrow's Impossibility Theorem. See Temkin, Intransitivity and the Mere Addition Paradox, 159.

[27] A. Tversky and D. Kahneman (1987) Rational choice and the framing of decisions, *Journal of Business*, 59, 5251–5278. If external alternatives did not affect preferability through such considerations as regret, psychologists would observe transitivity in decision-making, but they do not.

[28] A parallel person is a person who would have been the recipient of the same interests that you represent — who would have existed 'in your place', given the realization of the other alternative. See Kai M. A. Chan, Duties to future persons.

[29] Tversky and Kahneman, Rational choice and the framing of decisions.

[30] For more on parallel people, see Chan, 'Duties to future persons'.

[31] See Kai M. A. Chan, (submitted) The Golden Rule and the potentiality principle: future persons and contingent interests, *Analysis*; and Chan, Duties to future persons.