Ekonomia — Wroclaw Economic Review 21/3 (2015)
Acta Universitatis Wratislaviensis
No 3712

Jakub Bożydar Wiśniewski
King's College London
jakub@cantab.net

Repugnancy, Marginalism, Transitivity, and Population Ethics

JEL Classification: I3, O1

Keywords: Repugnant Conclusion, population ethics, axiom of transitivity, diminishing marginal utility

Abstract
Repugnancy, Marginalism, Transitivity, and Population Ethics

This paper attempts to build upon the "marginalist" solutions to various puzzles in the area of population ethics, including the so-called Repugnant Conclusion (seen as a major obstacle to the viability of the Total Utility Principle), the Ecstatic Psychopath Scenario (seen as a major obstacle to the viability of the Average Utility Principle) and the Negative Repugnant Conclusion. After rejecting the suggestion that the above puzzles should be resolved by abandoning the axiom of transitivity, I argue that the solution lies in the principle of diminishing marginal utility, whose effects apply not only to every individual added to any given population, but, even more importantly, also to the already existing members of that population.

1. Introduction

The question for this paper is whether certain puzzles regarding interpersonal value conflicts, as well as certain highly counterintuitive conclusions resulting from these puzzles, impugn the power of our rational faculties to think coherently about the value of states of affairs with variable quantities and qualities of interpersonally conceived well-being. In other words: what are the prospects for inventing rational methods of making optimal choices in situations involving interpersonal value conflicts and value trade-offs?

To be clear, the question considered in this paper will not be the one of how and whether at all interpersonal comparisons of well-being (which term I shall use interchangeably with "welfare" and "utility") can be made. This is a very im-

portant and weighty problem in itself, but I believe that it lends itself more to the exploration of our empirical rather than rational capacities — to the exploration of the extent to which we can successfully empathize with others (or "put ourselves in others' shoes"), not of the extent to which we can accommodate intellectually the results of such empathy. This is perhaps the area most fruitfully investigated by the modern philosophy of mind, most notably by various "simulationist" and "co-cognitive" approaches (e.g., Goldman 1989, Gordon 1986). And even though I think that such investigations are highly unlikely to establish the existence of any relevant *cardinal* scale of measurement, I believe that they might lend some psychological credence and clarity to the processes whereby we make *ordinal* interpersonal comparisons of utility. The fact that people often allow other people to decide on their behalf and feel benefited by such life policy (think of, e.g., disciples vis-à-vis their teacher, employees vis-à-vis their boss, etc.) seems to be a fair indicator that at least some of us are indeed capable of making such interpersonal comparisons with a relative degree of success, and further research into the theory of mental simulation might illuminate the details of these capabilities.

However, as I said earlier, even if we were absolutely certain about the viability and extent of our empathetic skills, a still more pressing problem would remain, namely how to deploy these skills in order to achieve morally optimal results. Should one pursue the policy of maximizing total well-being, and if so, should one strive to create additional satisfied human beings or restrict oneself to increasing the welfare of the existing ones? In either case, where should one stop, if at all? Or perhaps one should strive to maximize average rather than total well-being? Perhaps the question of which value should be promoted depends on the state of the world? I do not intend to provide definitive answers to these and similar puzzles, clustered in the area known under the name of population ethics, but rather to defend the view that we can think about such questions coherently without reaching seemingly contradictory and intellectually implausible conclusions. The fact that, as we shall see shortly, such conclusions do indeed seem to follow from the above-mentioned inquiries appears to constitute a formidable obstacle that a viable rationalist approach to the issue of interpersonal value conflicts needs to overcome.

Here it is worthwhile to address the question of the range of moral theories that my remarks are supposed to apply to. An intuitive thought might be that given my focus on interpersonal utility comparisons, the issues that I shall discuss are pertinent only to utilitarianism. I do not think that this is true. It seems to me that every serious moral doctrine should aim at specifying methods for bringing about morally best results, and thus every moral doctrine might be tempted to deploy its own version of interpersonal utility calculus. This is a view shared by, e.g., Scanlon (1975: 655) and Rawls (1971: 26). I take it that just as for classical utilitarians the best consequences result from implementing the single principle of maximum total agent-neutral happiness, for Kantian deontologists they result from acting on

the Categorical Imperative, whereas for Aristotelian virtue-ethicists — from "living in a manner that actively expresses excellence of character or virtue" (Haybron 2000: 210). Admittedly, utilitarianism (as well as various forms of contemporary consequentialism) might be the only one of the above that is *exclusively* result-oriented, but it seems hard to deny that the others contain substantial teleological elements as well. Moreover, some adherents of not-fully-teleological moral doctrines may be inclined to perform utility comparisons of states of affairs in which (interpersonally quantified) utility of consequences is in some sense weighed against utility of intentions. Of course, to make the matter fully clear more would have to be written about the way in which various conflicting positions in moral theory understand result-orientedness and determine the contributory values of results. I believe, however, that my brief remarks in this paragraph suffice to indicate that there are no clear-cut and obvious restrictions on the range of moral theories that the problems considered in this paper are relevant to.

Having said the above, let us now investigate the puzzles associated with population ethics more closely.

2. The Repugnant Conclusion

Perhaps the most notorious scenario in which our intuitively positive appraisal of one value (the total well-being of a huge population) clashes with our intuitively negative appraisal of another value (the well-being of each member of that population) is the one described by the so-called Repugnant Conclusion (hereafter RC). RC states the following: "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living" (Parfit 1984: 388).

There are several ways of arriving at the RC, but it should suffice for the present purposes to describe just one of them, the one that I deem to be the most straightforward and compelling (taken from Rachels 2004). Let A be a population of ten billion lives of very high quality and let B be a population ten times as big and enjoying the quality of life almost as high, but still somewhat lower than that enjoyed by the members of A. Given a hugely plausible claim that a small decrease in quality can be outweighed by a great increase in quantity, one should conclude that B is (morally) better than A. But then we can iterate the process of sacrificing small amounts of quality at the expense of gaining great amounts of quantity until we reach an immensely huge population (Z) whose members live lives barely worth living. Now, given the transitivity of the relation of betterness, we seem bound to accept that Z is better than populations such as A or B, which is deeply counterintuitive. However, unless we can find some way to block the RC

or explain away its counterintuitiveness, it appears that we have to acknowledge that practical reason cannot deal adequately with interpersonal value conflicts (in this case the conflict between the quality and quantity of well-being) as related to the global scale.

3. Solving the problem, attempt one: A doubt about transitivity

A number of solutions to dealing with the RC have been proposed over the last three decades. Space constraints do not allow me to do justice to all of them — I shall try to elaborate and build upon the ones that seem to me to be the most promising. Let me start, however, from focusing on the one that appears the most revisionary with regard to the perceived nature of practical reason. More specifically, I am referring to Larry Temkin's contention that rationality might not require maintaining a transitive order of rank on one's preference scale (Temkin 1996).

Temkin builds his "continuum argument" for intransitivity on the basis of the following three claims:

1) For any unpleasant or "negative" experience, no matter what the intensity and duration of that experience, it would be better to have that experience than one that was only a little less intense but twice as long.

2) There is a continuum of unpleasant or "negative" experiences ranging in intensity, for example, from extreme forms of torture to the mild discomfort of a hangnail.

3) A mild discomfort for the duration of one's life would be preferable to two years of excruciating torture, no matter the length of one's life (ibid.: 179).

Taken together, these apparently plausible statements are supposed to challenge the axiom of transitivity and thereby also block the RC. Temkin's argument in a nutshell could be stated thus: let A be a life that contains two years of intense torture and let B be a life that contains four years of torture almost as intense, but still somewhat weaker than that suffered in A. Again, it seems uncontroversial to suppose that a small decrease in intensity can be outweighed by a considerable increase in duration, and that therefore B is worse than A. And again, by iterating the process of exchanging small amounts of intensity for great amounts of duration we will finally reach an immensely long life (Z) of a very mild discomfort (e.g., an annoying hangnail). Given the transitivity of the relation of betterness, we now have to infer that Z is worse than A and B, but such a conclusion is as implausible as the RC. Thus, by constructing a thought experiment structurally parallel to that leading to the RC, but even more disagreeable with our intuitions, Temkin argues for the rejection of the axiom of transitivity.

However, the consequences of such a rejection seem no more palatable than those that Temkin's argument tried to save us from. Let us see how they affect the

scope of practical reasoning. Suppose that A is better than B, B is better than C and C is better than A. So which of these should I buy as a present for my friend? Obviously enough, I want to buy the best item available — but which is best? Certainly not A, since C is better. Certainly not C, since B is better. And certainly not B, since A is better. As it turns out, given intransitivity, our capacity for optimizing choices is completely stunted. Temkin himself seems very much aware of such problems when he says:

> Reflection suggests that most alternatives might be placed on a large, intransitive continuum, analogous to the one involving extreme torture at one end and mild discomfort at the other. This opens the possibility that there would be no rational basis for choosing between virtually any alternatives. (ibid.: 209)

But does the continuum argument really undermine transitivity? Apparently, Temkin's rationale for claiming that it does is his contention that "together a sufficient number of differences in degree can sometimes amount to a difference in kind" (ibid.: 194). This is ostensibly illustrated by the fact that no amount of hangnail can possibly outweigh or even equal two years of excruciating torture. I personally find Temkin's contention about degrees and kinds quite plausible, but I do not think that it implies a refutation of transitivity. After all, transitive relations operate on lists of items that can be subsumed under a *single* yardstick of comparison, ranging from fairly general (e.g., prudential or moral betterness) to quite specific (e.g., efficiency in combating cold). And even though, admittedly, incorporating multiple factors into such a yardstick might complicate the task of ranking a given list of outcomes, it does not undermine the conceptual fundamentals of the very process of ranking (whereas rejecting transitivity would constitute precisely such an undermining).

Think of a mundane case such as eating apples. It appears natural to suppose that it is better for a connoisseur of apples to eat two apples rather than to eat just one. Likewise, it might be even better for him to eat three apples, and so forth. But we cannot iterate this train of thought indefinitely. As soon as the state of surfeit is reached, eating every additional apple will make things worse for the person in question. Thus, if X reaches the state of surfeit after eating, say, seven apples, it is true that he would prefer eating five apples to eating four apples and that he would prefer eating six apples to eating five apples, but it is not true that he would prefer eating seven apples to eating four apples. Is this another example of intransitivity? It seems clear to me that it is not. The point is that by moving from the sixth apple to the seventh apple we moved from the category of outcomes that generate nutritive satisfaction to the category of outcomes that generate nutritive excess. And, as I mentioned before, transitivity operates only within, not across distinct categories (or kinds) of outcomes.

Note that the above principle applies not only to outcomes associated with obtaining specific collections of items, but also to outcomes associated with obtaining individual items from these collections. Every additional item of a single

kind produces lower utility for its user due to the law of diminishing marginal utility. Thus, the second apple represents a lower subjective use-value than the first apple. Likewise, obtaining the second item in a collection produces less utility than obtaining the first item. However, obtaining the item that completes the collection (say, the 100th item) may produce much more utility not only than obtaining the 99th item, but possibly also than obtaining all previous 99 items taken together. This is because completing the collection marks the crossing of a utility threshold (Frankfurt 1987: 27–28). In other words, it marks the move from one category (items constitutive of the collection X) to another category (collections of the kind X). Such a move is consistent both with the axiom of transitivity and with the law of diminishing marginal utility (completing the second collection of the kind X would produce less utility than completing the first, unless bringing two collections of the kind X together would produce a generically different *super*collection of the kind, say, Y, which would mark the crossing of yet another utility threshold).

In view of the above, we can now see that Temkin's continuum of discomfort is not really a continuum, since it includes at least one threshold of disutility — namely, the threshold dividing the category of unbearable discomforts (e.g., excruciating torture) and the category of bearable discomforts (e.g., mild hangnail). The axiom of transitivity operates only within each of these categories, but not across them. And yet, Temkin's contention is, presumably, that we have a good reason to suppose that intransitivity can occur within a single category of items. By now it should be clear that in fact no such reason follows from his arguments. And even though he admittedly anticipates a counterargument similar to mine when he discusses the possibility of claim 1 no longer holding "once an unpleasant experience becomes mild enough" (Temkin 1996: 182), he apparently does not take account of the crucial fact that such an experience belongs to another continuum (i.e., the continuum of bearable discomforts), rather than to the one hitherto discussed (i.e., the continuum of unbearable discomforts).

Thus, the feeling that Temkin's thought experiment establishes the intransitivity of the relation "better than" and hence challenges the nature of practical reasoning should now be put to rest. However, the price to be paid for preserving the standard shape of rationality is that we once again become open to the threats of the RC. That said, I believe that some observations made in this section of the paper shall prove useful in our subsequent search for a satisfactory solution.

4. Solving the problem, attempt two: Total or average well-being?

Another way to block the RC is to suggest that it is not total but *average* well-being that ought to be promoted. Since the average welfare in the overpopulated

repugnant[1] world is very low, such a world clearly does not meet the criteria of a good world according to the theory advocating the maximization of average welfare (from now on, I shall designate it by the term "average theory", as opposed to "total theory").

But the task at hand would be too easy if the average theory did not have its own problems, i.e., if it were not susceptible to its own version of the RC. Consider the case of the Ecstatic Psychopath (hereafter EP): EP is the person with the highest level of welfare in the world. The presence in the population of any person whose well-being is lower than that of the EP (note that in the case under consideration that includes *every* person) will consequently decrease the average well-being of that population. Hence, one is bound to conclude that according to the average theory the optimal population consists of the EP only, which seems pretty disagreeable, if not as repugnant as the original RC. An adequate description of such a scenario might require filling in some details, for instance insisting that the EP's psychopathy precludes any welfare-generating cooperation with other human beings (who could otherwise possibly become as happy as the EP is) and that the rest of the population is to vanish peacefully, e.g. due to being persuaded by the EP not to reproduce (since a brutal extermination could cause enough suffering to outweigh the subsequent well-being of the EP). But even without such ceteris paribus qualifications the EP scenario appears to spell some serious trouble for the average theory.

In sum, both the total and the average theory seem to have something to be said in their favor, but neither seems capable of avoiding generating certain highly counterintuitive implications. Could the solution lie in combining the best elements of each? A proposal along these lines appears to be proposed by Thomas Hurka (1983), who sketches what he calls the "variable value view". Adopting this view enjoins us to follow the total theory with regard to small populations and the average theory with regard to overcrowded populations. The rationale behind Hurka's suggestion is the principle whereby the utility contributed by an additional happy person is a diminishing function of the total population that he inhabits. In other words, happy persons are supposed to exhibit diminishing marginal utility vis-à-vis the world they occupy.

I am not sure as to Hurka's preferred explanation of the phenomenon in question — he mentions our perceived duty to care for the preservation and flourishing of the human species, which supposedly becomes less stringent as the human population grows, but he does not elaborate on these remarks very much. Let me

[1] As Rachels (2004) rightly points out, what is supposedly repugnant with regard to the RC is not the overpopulated world whose inhabitants live the lives barely worth living, but the conclusion that such a world is better than that inhabited by a much smaller population enjoying a much higher standard of living. However, for the sake of simplicity I shall henceforth use the adjective "repugnant" to designate both the disturbing conclusion in question and the conditions of the world that it describes.

therefore sketch my reconstruction of the kind of reasoning that could lead to embracing the principle of the diminishing marginal utility of happy persons.

The RC results, among others, from relying on a simplified, hypothetical model of cardinally measurable and temporally homogeneous utility. Hence the talk about each additional person contributing *x* units of well-being to the population, where *x* is a cardinal number. But in fact utility has neither of these characteristics; it is, to a substantial degree, a mental or psychological magnitude, and, unlike physical magnitudes, it cannot be measured against any universal, spatially extended, physical scale. It can be measured ordinally — most certainly against each person's individual, psychological preference scale, and perhaps, though it can never be told in advance with what degree of accuracy, against the preference scales of others. Nevertheless, be that as it may with the predictive precision of any single of such ordinal interpersonal comparisons, all of them reveal the universal validity of the law of diminishing marginal utility.

The law in question follows from the fact that humans are purposive beings and that the actions they employ for the satisfaction of their purposes are always extended in time. Thus, these purposes cannot be all accomplished simultaneously — they have to be arranged according to some ranking of importance, where the earliest actions are aimed at the satisfaction of one's most central aims. The first glass of water can save one from dying of thirst, the second can restore the strength of one's organism, but the millionth is practically of no value — a statistical person simply does not have that many aims requiring the use of water, and its utility diminishes towards zero as more and more of these aims are satisfied. Now, perhaps Hurka's intention is to look at the human population as an individual human being writ large. If the population consists of only two members (of both sexes), then their foremost duty is to procreate and thus preserve the species (ibid.: 497). However, as the survival of the species becomes more and more secure, further population increases contribute less and less utility, and the principal duty of the "collective organism" gradually shifts to that of increasing its average well-being.

Now, having explained the possible rationale behind the variable value view, we need to ask what are the problems that it has to countenance. Some of these have been identified in a paper by James Hudson (1987). For in- stance, it seems that the variable value view involves a certain arbitrariness with regard to specifying the cut-off point beyond which the total theory is to give way to the average theory and vice versa. Here is the relevant quotation from Hudson's paper:

> For a complete statement of the theory must tell us, as Hurka fails to do, what constitutes a "small" population and what constitutes a "large" one, and just how the value of an extra happy person falls as population increases. (1987: 132)

Furthermore, it appears that the variable value view is incapable of shaking off the weaknesses of the average theory — it is hard to see, for example, how it can avoid admitting that one of the average-welfare-increasing methods of deal-

ing with the RC could consist in killing off[2] relatively unhappy members of the population. Analyzed purely in terms of its conduciveness to augmenting average utility, such a method seems as effective as that of holding a tight rein on the population's growth.

Nevertheless, I believe that Hurka's proposal is on the right tracks and that some sort of marginalist theory is capable of dealing adequately with the problems facing both the total and the average view. Let us now explore this possibility further.

5. Solving the problem, attempt three: Another thought on marginalism

What seems to me problematic about Hurka's suggestion is that he believes that the principle of diminishing marginal value of happy persons holds unconditionally — that is, regardless of the effects that increases in the population will have on already existing people. In the author's words:

> I have argued that, even when population increases will have no effect on the average well-being, we think they are more important at low levels than at high levels, and that the average and total principles cannot capture this view. (Hurka 1983: 499)

He considers the effects in question as *side effects*, and finds them irrelevant to the applicability of the marginalist principle that he espouses. Such an approach, however, appears to carry some bizarre metaphysical undertones — it genuinely treats the population as a separate, collective being with its own utility curve, a being which "collects" or "consumes" human beings by incorporating them into its organism, just as a collector gathers items into a collection (where acquiring every additional item produces less utility) or as a water-based organism replenishes its strength by drinking glasses of water (where drinking every additional glass produces less utility). My intuition, on the contrary, is that the fact that at some level population increases start to produce less and less utility stems precisely from the effects that they have on already existing persons (which, in turn, determine the well-being of "newly added" people, so the influence here is mutual).

Before elaborating on the above remark, let me introduce what is in my opinion a crucial caveat: if we are to take seriously the role played by the phenomenon of diminishing marginal utility, then the scenarios we analyze cannot be "static" or "punctual"; if they are to retain relevance to the real world, they have to be extended in time. In other words, what should interest us is determining the conditions characteristic of "optimal lives" rather than "optimal spatiotemporal life-

[2] Perhaps one should add the qualification "suddenly and painlessly" in order to stipulate away possibly great amounts of disutility felt by those exterminated, amounts perhaps sufficient to actually bring the average level of well-being down rather than up.

slices". Thus, for instance, it is quite implausible to suggest that the scenario in which the only member of the human population, an ecstatic psychopath, receives a pleasure shock resulting in 10,000 utils[3] and lasting one second (after which his life comes to an end), is a happy scenario. Moreover, in order to avoid embroiling ourselves with the potential problem of determining how to compare the value of scenarios where the same amount of well-being is spread over time intervals of different lengths and different temporal arrangements of the same set of events, let us assume that we shall always compare scenarios lasting the same amount of time and exhibiting regular patterns of ever-repeating series of happenings[4]. I do not think that such assumptions are in any relevant sense question-begging, most importantly because they are not discriminatory towards any of the considered theories, since neither of the analyzed "repugnant" scenarios (by which I mean and shall henceforth mean both the "overpopulation scenario" and the "ecstatic-psychopath-induced underpopulation scenario") assumes greater volatility or intertemporal uncertainty than the other.

In view of the above, the answer to the naturally following question "how long should the analyzed model scenarios be" should presumably be: long enough to matter for a realistically conceived human population; e.g., as long as the life of a single generation (i.e., 66 years, which is the current life expectancy of the world). However, given our auxiliary assumption of "regularity" (or "repeating history"), we could focus our analysis on considerably longer scenarios too, spanning the lives of up to several generations. The upper bound here seems to be our horizon of predictability, whereas the lower bound — as mentioned before — relevance to the lives of at least one generation.

Now let me move directly to the investigation of what I take to be the reasons for and the implications of the occurrence of the phenomenon of diminishing marginal utility as applied to the area of population ethics. As I argued contra Hurka, what I take to determine the value of adding new (at least somewhat) happy people to a given population is not its size, but the effects that such additions will have on the well-being enjoyed by the newly expanded population. Let me now explain what kinds of effects I have in mind. As new individuals are brought into the population, consumable goods, including space, become more and more

[3] As I said before, I am skeptical about the possibility of cashing out utility in cardinal terms. The cardinal numbers used throughout this chapter are to be understood exclusively as a useful mental shortcut, employed in order to make my examples more vivid and exact.

[4] To explain what I mean by "regularity" on the basis of examples: in the overpopulation case, it simply assumes constant growth of the population and no significant technological changes with respect to the possibilities of accommodating such growth; in the "ecstatic psychopath" case, it assumes that the world is inhabited by a "dynasty" of psychopaths, so that at any given moment the sole member of the world population is a psychopath, who upon his death spawns another psychopath, and so on ad infinitum.

scarce[5]. The result of this increased scarcity is that every additional person will have fewer opportunities for enjoyment.

Of course, the division of labor and capital accumulation (as well as the resultant technological advancement) can outstrip and thereby mitigate or even completely overcome the effects of resource overutilization, but only given that the phenomena in question develop and proceed at a rate faster than that of the population growth. However, in all "repugnant" scenarios population growth eventually outstrips the above-mentioned mitigating factors by stipulation (which, in any event, is not an implausible stipulation: space, for instance, seems to be an inexorably scarce good).

Consequently, as the pool of rivalrous goods becomes depleted, more and more highly enjoyable things and amenities (e.g., convenient means of transport, technological gadgets, modern medicines, manifestations of the so-called "high culture" etc.) become permanently inaccessible[6]. Thus, it seems plausible to conclude that in the RC scenarios the utility of adding new human beings to the population is already so low that no number of new additions can compensate for the utility losses brought about by the disappearance of those highly enjoyable amenities whose production requires more resources per capita than is available under "repugnant" conditions.

One might claim that, again, it is still a matter of scale — given enough additions, an enormous multitude of almost-miserable lives will ultimately produce more total utility than a much smaller number of high-quality lives. This objection, however, misses the crucial point that I have tried to highlight — namely, the fact that from a certain point onwards such additions come at a serious price. As soon as the maximum carrying capacity of the planet is surpassed (and no available civilizational solutions are able to increase it), not only will every added person have less opportunities for enjoyment, but they will also decrease the range of opportunities available for those persons already in existence (due to inevitably rivalrous consumption of ineluctably scarce goods). Just as the sum of the terms of the infinite series 1/2+1/4+1/8+1/16… tends to 1, but can never equal it, the sum of the utilities of the persons brought into the world of resource overutilization can never equal (let alone exceed) that of the utilities of the persons who comprise a population small enough not to trigger the downward spiral leading to the RC.

[5] Except for the so-called "free goods", such as air and fresh water.

[6] One might claim, borrowing a piece of terminology from Parfit (1984), that deploring the loss of such high-flying amenities caused by bringing additional (still comparatively happy) human beings into the population is "elitist". But I see nothing elitist about it. As I understand what Parfit calls an elitist view, it is the view according to which we should attach the highest importance to maximizing the happiness of the most happy, not the view that we should not interfere with the happiness of the most happy or that we are not allowed to obstruct their opportunities for achieving even greater happiness.

Furthermore, under repugnant conditions everybody is essentially arrested in a dull, monotonous and extremely unrewarding life (all it can offer is metaphorically described by Parfit as "muzak and potatoes"); this monotony and lack of prospects for improvement is likely to aggravate the undesirable effects of the law of diminishing marginal utility, even if we put all considerations of rivalrous consumption and resource overutilization aside. In other words, if I get fed up with potatoes, the best way to prevent my "nutritional" utility curve from decreasing is to turn to a different kind of food. And the same goes for every other kind of activity. Such a solution, however, is impossible under repugnant conditions, while remaining a perfectly viable option in most other scenarios.

Of course, all of the above follows only if we are willing to take the phenomenon of scarcity seriously enough, but since the vision of the repugnant world is hardly compatible with the assumption of Edenic superabundance, I can safely presume that all philosophers who work on the problems of population ethics take scarcity as an undeniable and ineradicable fact.

How shall we apply the above observations to the dispute between the total and average theories? Do they offer some "third way" or some prospects for reconciliation, stronger than Hurka's? I believe that they do.

We need to note that the RC offends our moral intuitions not only because it leaves an average member of the population it describes with very little well-being. Its additional disagreeability, which is rarely noticed but which effectively removes its air of paradoxicality, derives from the fact that, if conceived of realistically, every RC scenario makes the total well-being of the described population very small as well.

Let us remember that the law of diminishing marginal utility not only brings it about that, as soon as the available resources start to become overutilized, the life of every additional human being is going to be of a lesser quality — it also brings it about that every such addition is going to decrease the quality of lives of those already in existence. Can the sum total of such meager utilities be big, especially given that the more elements it has, the less each of these elements amounts to? I think not. Hence, given a sufficiently long period of time (which need not be very long if the rate of population growth is great enough), the utility curve of any "repugnant" world will quickly approach zero. From then on, the lives lived by the members of any "repugnant" population, measured over any time interval, will consistently produce a total utility close to zero. Therefore, even if in any given scenario the years preceding the beginning of the process of resource overutilization are very conducive to the production of well-being, a sufficient number of "repugnant" years are easily capable of outweighing the blessings of their predecessors. But in the world in which, say, during a century, there are no "repugnant" years (due to the fact that the population is smaller, resources are not overutilized and thus each inhabitant is able to enjoy a range of amenities rich and satisfying enough to produce quantities of utility even a fraction of which

cannot be attained in any "repugnant" world), its population will enjoy, within that century, not only a higher average level of utility, but also a higher total level of utility.

Thus, it is not really the case that there is some critical mass point beyond which one should abandon the total theory in favor of the average theory — in view of the above observations, it is likely that we need not help ourselves to this distinction at all in order to deal with the RC. In other words, the repugnant scenarios are repugnant on both scores — with regard to total well-being as well as with regard to average well-being, although the former seems not to be realized very often.

Incidentally, the same conclusions apply to the Ecstatic Psychopath scenario, which shows that the reduction of total welfare goes hand in hand with the reduction of average welfare in a variety of repugnant cases. Let us quickly recall that the troublesome character of the EP scenario involves *under*utilization rather than overutilization of available resources, i.e., the psychopath blocks the existence of potentially very many happy and accomplished human beings. The most extreme variety among the possible EP scenarios makes its hero not only a psychopath, but also a Nozickian "utility monster" (Nozick 1974: 41), i.e., an entity that is immensely proficient at producing utility from whatever enjoyable action it engages in. As a result, the problem is not only that the presence of others will decrease his well-being; more importantly, the problem is that no matter how many normal, happy people will be brought into existence, their happiness will be unlikely to outweigh his lost ecstasy. Is it therefore untrue that losses in total utility must ultimately converge with losses in average utility?

An easy answer to the above-mentioned worry is to invoke the postulate of psychological plausibility — presumably, what makes the original RC repugnant is, among others, the fact that we do not think of the population involved as composed of hermits or ascetics who would be quite content with living materially very destitute lives. This is psychologically plausible, consistent with what we know about the majority of humankind. But then in the EP scenario we stipulate the existence of a psychologically implausible being, whose sensational capabilities are unlike anything observed in reality.

Such a response might seem *too* easy and could be countered with the claim that it is not statistically implausible for the ecstatic psychopath to be a singular anomaly, but it is much more implausible to suggest that the majority of any given human population could turn into hermits or ascetics. Hence, the EP scenario should not be thought of as psychologically improbable.

There is, however, another kind of psychological, or perhaps even conceptual implausibility that the assumptions underlying the EP case can be accused of. More specifically, there is no reason to think that the ecstatic psychopath can be exempt from being affected by the law of diminishing marginal utility. As a purposeful, intentional being, which cannot pursue, let alone accomplish all of its

aims simultaneously, the psychopath, like all of us, must act according to a specified ranking of desires, starting from the actions aimed at satisfying those ranked most important and then gradually moving towards the pursuit of the less and less pivotal. This principle holds even if one entertains a single desire stretched over a long period of time — the earlier moments of its satisfaction are more conducive towards one's well-being than the later ones. I do not see a good reason to think that the same does not go even for everlasting ecstasy — it seems fair to believe that no matter how unbelievably intense it might initially be, given enough time it is bound to become just as dull and unrewarding as any endlessly repetitive phenomenon.

Consequently, given enough time, the yearly average utility enjoyed by the single-member population from the EP scenario is going to drop below 1 and tend towards 0.

One might protest that here we are relying on a simplified understanding of utility as something produced by a single, homogeneous kind of sensation or experience, perhaps somewhat akin to drug-induced pleasure, whereas the correct way to think about the psychopath is that he reacts ecstatically to a variety of experiences (associated, e.g., with interacting with nature or pursuing creative efforts) and continually looks for new ones. Thus, it is wrong to think that his interaction with the world on the whole will ultimately become as dull as drinking another glass of water.

There appears to be a degree of truth to this objection, but I do not think that even this "revised" image of the psychopath's life can escape the strictures of the law of diminishing marginal utility. There are, after all, a finite number of general classes of activities that we can undertake, and which can be described with varying, oftentimes overlapping, degrees of detail — e.g., creative work, exotic travel, food tasting, writing a poem, climbing a mountain, etc. Whatever a person decides to do falls into one or more of these classes — visiting a new, beautiful place is, by definition, a new experience, but it is nonetheless an old *type* of experience insofar as it involves visiting just another picturesque spot. An avid globetrotter might be quite impressed by a hundredth palace he has seen in his life, but I believe it unlikely that, regardless of its beauty, he might find that experience as intensely enjoyable as the one he felt upon seeing the first one. The same seems to go for every other hobby one could have, as well as for various mixes of hobbies.

Thus, even if it is incorrect to think of the psychopath's utility curve going steadily downward towards zero, a psychologically plausible forecast is that the law of diminishing marginal utility will bring it down to and stabilize it at the level of a statistically normal human being who is enthusiastic about his hobbies. To think that the psychopath could remain genuinely ecstatic about all his undertakings throughout his entire life, appears to me, as I argued, implausible.

In other words, given enough time, the yearly average utility enjoyed by even the most multi-faceted and versatile psychopath will, after falling for some time,

very likely become a more or less constant value. It seems reasonable that at that stage the only way to bring about a substantial increase in it would be to infuse the world under consideration with, quite literally, new blood — i.e., new human beings, not yet aware of all the pleasurable and rewarding experiences that life can offer. And even if the utility curves of these new people were eventually to stabilize as well, their sheer number and the possibility of them bringing even more potentially happy persons into existence is, I think, certainly enough to outweigh the disutility that the appearance of other human beings might cause to the psychopath. In conclusion, the EP scenario, which is supposed to cause trouble to the supporters of the average theory, is likely to contain less average utility than non-EP scenarios, just as the RC scenario, which is supposed to cause trouble to the supporters of the total theory, is likely to contain less total utility than non-RC scenarios.

Let me end this section by giving a moment's considerations to the question of whether the marginalist insights I drew upon might also help us to overcome what is known as the Negative Repugnant Conclusion (Broome 2004: 213), which says that a very large population whose members live slightly unhappy lives (hereafter P1) is worse off than a substantially smaller population whose members live horrendously unhappy lives (hereafter P2).

I believe that the answer is yes. Let us notice that precisely because the phenomenon of diminishing marginal utility is true, there is no corresponding phenomenon of diminishing marginal disutility — the first draught of water is so precious to a thirsty person precisely because it helps to liquidate an immense disutility associated with being thirsty. Likewise, eating the second piece of cake is less satisfying than eating the first piece of cake, but it is definitely not the case that passing up the second offered piece of cake makes one's hunger less vexing than passing up the first offered piece of cake. Finally, every additional second of torture is harder to stand, up to the breaking point, but it is not the case that every additional second of pleasure is more pleasant.

Thus, given that disutility is marginally increasing, it seems plausible to suppose that it would be very difficult for the total disutility of a very large population living just below the level worth living to offset the disutility of a smaller population whose members experience immense suffering. An objector might insist that given sufficiently many members in the former population (P1), the negative repugnant conclusion will still follow. But there is one important blocking factor for this kind of reasoning. Let us remember about the inevitable phenomenon of scarcity — in order to preserve the credibility of the scenario under consideration, we cannot add new members to P1 indefinitely, since from a certain point onwards introducing every additional member into it will ultimately lower the marginal productivity of the available resources, and thus further increase the disutility of each of its members. Hence, it appears unavoidable that over time their scarcity-induced suffering (caused by, e.g., hunger, lack of shelter, lack of sanitation

and consequent health risks, etc.) will reach the level of suffering of the members of P2, and since P1 was supposed to be larger from the outset, its total disutility is ultimately more than likely not only to equal, but to exceed that of P2. This result, however, should come across as only too natural and not repugnant at all.

Another objection could consist in questioning that utility is in fact marginally increasing by citing the fact of habituation. After all, is it not true that a permanently hungry person could possibly be less troubled by their condition that a person who experiences real hunger for the first time in their life? Such a suggestion, reasonable as it is, seems to me to confuse two distinct phenomena. To describe the issue in economic terms, habituation appears to involve not so much a change in the shape of one's disutility curve, but rather a simple shift in it. Admittedly, exposure can move the threshold at which one finds a given bad discomforting, but that does not invalidate the principle that beyond the threshold in question more of the bad is more discomforting. And even though it is true that exposure to some bads is capable of bringing about such a threshold shift, exposure to other bads (including the ones afflicting the members of P2) probably cannot.

Perhaps the above-mentioned two categories could be roughly described as mere disutilities and torments. My contention is that while mere disutilities of one's life can be outweighed by utilities, which makes the overall disutility of a generic "slightly unhappy life" marginally decreasing, genuine torments cannot be outweighed by any amount of attendant utility (in other words, unless a genuine torment ceases, no otherwise happy event can be enjoyed by the tormented person), thus making the overall disutility of a "horrendously unhappy life" marginally increasing. This conclusion lends some justification to the Temkin-style claim that no amount of hangnail pain can outweigh two years of extreme torture (Temkin 1996: 194). After all, it is hard to deny that life with a permanent hangnail might still have some (or even a lot of) intrinsic value, opportunities to choose from, goals to achieve, desires to act on and satisfy, etc. In short, it is not putting a person in the state of incapacitating pain. On the other hand, it is, I believe, possible to argue that a life of torture, or even a two-year period of torture (provided that it cannot be alleviated), has no intrinsic value, but only a huge intrinsic disvalue.

Hence, it seems to me that the marginalist thinking employed in this paper is able to defuse both the Negative Repugnant Conclusion as well as all of its aforementioned variants.

Let us remember that the issues investigated here do not cover the question of what particular value should be promoted by reason in any specific situation or whether there is a single value that it is rational to maximize in every possible situation — I am inclined to believe that whether one should promote, e.g., total or average utility, is a decision that should take into account the specific circumstances of the scenario at hand, among which there might be, for instance, the consideration of whether or not there are huge disparities in well-being within the group of persons towards which we wish to act. The goal of this paper was

to investigate whether large-scale ethical questions involving interpersonal comparisons of utility can be addressed by practical reason without ending up with a variety of highly counterintuitive results, and I believe that the said investigation yielded a positive answer.

References

Broome, J. (2004), *Weighing Lives*, Oxford: Clarendon Press.
Frankfurt, H. (1987), 'Equality as a Moral Ideal', *Ethics*, 98, 21–42.
Goldman, A. (1989), 'Interpretation Psychologized', *Mind and Language*, 4, 161–185; reprinted in M. Davies and T. Stone (eds., 1995), *Folk Psychology: The Theory of Mind Debate*, Oxford: Blackwell Publishers.
Gordon, R. (1986), 'Folk Psychology as Simulation', *Mind and Language*, 1, 158–171; reprinted in *ut sup*.
Haybron, D.M. (2000), 'Two Philosophical Problems in the Study of Happiness', *The Journal of Happiness Studies*, 1, 207–225.
Hudson, J.L. (1987), 'The Diminishing Marginal Value of Happy People', *Philosophical Studies*, 51, 123–137.
Hurka, T. (1983), 'Value and Population Size', *Ethics*, 93, 496–507.
Nozick, R. (1974), *Anarchy, State, and Utopia*, New York: Basic Books.
Parfit, D. (1984), *Reasons and Persons*, Oxford: Oxford University Press.
Rachels, S. (2004), 'Repugnance or Intransitivity: A Repugnant but Forced Choice', in: J. Ryberg and T. Tännsjö (eds.), *The Repugnant Conclusion. Essays on Population Ethics*, Dordrecht: Kluwer Academic Publishers.
Rawls, J. (1971), *A Theory of Justice*, Cambridge, Mass: Harvard University Press.
Scanlon, T. (1975), 'Preference and Urgency', *Journal of Philosophy*, 72, 655–669.
Temkin, L.S. (1996), 'A Continuum Argument for Intransitivity', *Philosophy and Public Affairs*, 25, 175–210.