A Pareto Principle for Possible People

Preferences

Edited by
Christoph Fehige and Ulla Wessels

Walter de Gruyter · Berlin · New York
1998

♾ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Preferences / edited by Christoph Fehige and Ulla Wessels.
p. cm. – (Perspektiven der analytischen Philosophie ; Bd. 19 = Perspectives in analytical philosophy)
Based on a conference held in June 1992 in Saarbrücken, Germany, and in Saarlouis, Germany.
Includes bibliographical references and indexes.
ISBN 3-11-015007-7
1. Preferences (Philosophy) – Congresses. I. Fehige, Christoph. II. Wessels, Ulla. III. Series: Perspectives in analytical philosophy ; Bd. 19.
B105.P62P74 1998
128′.3–dc21 97-28699
CIP

Die Deutsche Bibliothek – Cataloging-in-Publication Data

Preferences / ed. by Christoph Fehige and Ulla Wessels. – Berlin ; New York : de Gruyter, 1998
(Perspektiven der analytischen Philosophie ; Bd. 19)
ISBN 3-11-015007-7 brosch.
ISBN 3-11-015910-4 Geb.

© Copyright 1998 by Walter de Gruyter GmbH & Co., D-10785 Berlin
All rights reserved, including those of translation into foreign languages. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without permission in writing from the publisher.

Printed in Germany
Typesetting: Thomas Fehige, Münster
Printing: Arthur Collignon GmbH, Berlin
Binding: Lüderitz & Bauer, Berlin
Cover design: Rudolf Hübler, Berlin

All I want is to sit on my arse and fart and think of Dante.

Samuel Beckett

Desires and wants, however intense, are not by themselves reasons in matters of justice. The fact that we have a compelling desire does not argue for the propriety of its satisfaction any more than the strength of a conviction argues for its truth.

John Rawls

Take any demand, however slight, which any creature, however weak, may make. Ought it not, for its own sole sake, to be satisfied? If not, prove why not. The only possible kind of proof you could adduce would be the exhibition of another creature who should make a demand that ran the other way. The only possible reason there can be why any phenomenon ought to exist is that such a phenomenon actually is desired.

William James

Preface

Preferences is a collection of essays on the concept and the role of preferences (desires, and the like) in practical reasoning. Ground covered includes welfare, prudence, rational decision making, and all areas of moral philosophy: ethics (applied and not so applied), metaethics, and deontic logic. A special symposium looks at *possible* preferences and their significance in matters of life and death, including the notoriously thorny question how many people there should be. All the essays are published here for the first time.

The book is not just for specialists. We have given it an introduction that, though it may move swiftly, at least starts from scratch; a selected bibliography is also provided.

Most of the authors were able to meet in advance, and to present, discuss, and then revise their contributions. But the line has to be drawn somewhere, and authors who receive a reply in this volume were not permitted to adjust their papers in the light of the final version of the reply. The initial exchange took place in Saarbrücken and Saarlouis in June 1992.

*

Everybody has been very kind to us. Georg Meggle – selfless and cheerful as usual – co-designed the project and supported it from beginning to end. When we proposed the meeting, we were backed up by Franz von Kutschera and Wolfgang Lenzen. Barbara Schumacher helped prepare and run it.

The editors of *Perspectives in Analytical Philosophy*, Georg Meggle and Julian Nida-Rümelin, have welcomed the book in precisely the form we suggested. The authors have been co-operative and patient throughout. Christopher Abbey and Seán Matthews have given valuable advice, linguistic and otherwise, to many of us. Kornelius Bamberger was able, and kind enough, to convert most of the data that the contributors sent us. Thomas Fehige gave these data a neat, uniform lay-out. Patrick Agsten, Monika Claßen, Franziska Muschiol, Ulf Schwarz, and Valentin Wagner have assisted us, efficiently and in numerous respects; the same holds true of Karin Thom. With this list in chronological order, one important acknowledgement comes last: de Gruyter publishers. Working with Hans-Robert Cram was a pleasure; ditto, at the technical end, with Grit Müller.

The conference that gave rise to this book was made possible by the financial assistance of: Deutsche Forschungsgemeinschaft (DFG), Fritz Thyssen Stiftung, Ministerium für Wissenschaft und Kultur des Saarlandes, Universität des Saarlandes, and Vereinigung der Freunde der Universität des Saarlandes. The DFG (research project "Was zählt?") has also funded our own work on this volume.

We thank all these persons and institutions for their support.

*

We share the belief, now regarded in some quarters as both unsound and old-fashioned, that, in essence, morality is all about welfare, and welfare all about preferences. Some of the contributors to this volume would agree, some would not. With luck, this collection will help advance matters a little.

Christoph Fehige and Ulla Wessels
Leipzig, January 1998

Contents

Abstracts xiii

Introduction

CHRISTOPH FEHIGE AND ULLA WESSELS
Preferences – an Introduction xx
Preferences – a Short Bibliography xliv

Part I: Preference and Decision

ELIJAH MILLGRAM
Deciding to Desire 3
SYDNEY SHOEMAKER
Desiring at Will (and at Pill): A Reply to Millgram 26

CHRISTOPH LUMER
Which Preferences Shall Be the Basis of Rational Decision? 33
ANTONELLA CORRADINI
Intrinsic Desirabilities: A Reply to Lumer 57

RICHARD B. BRANDT
The Rational Criticism of Preferences 63
ANNA KUSSER
Rational by Shock: A Reply to Brandt 78

J. DAVID VELLEMAN
Is Motivation Internal to Value? 88
GEORG MEGGLE
Motivation and Value: A Reply to Velleman 103

SVEN DANIELSSON
Numerical Representations of Value-Orderings: Some Basic Problems 114
ULRICH NORTMANN
Interval Orders Defended: A Reply to Danielsson 123

Part II: Preference and Metaethics

LENNART ÅQVIST
Prima Facie Obligations in Deontic Logic: A Chisholmian Analysis Based on Normative Preference Structures 135
UWE BOMBOSCH
The Meaning of "Ought, Prima Facie" and Decision Situations: A Reply to Åqvist 156

FRANZ VON KUTSCHERA
Values and Duties 163
WILFRIED HINSCH
Beyond Duty: A Reply to von Kutschera 172

MARCUS WILLASCHEK
Agency, Autonomy and Moral Obligation 176
HILARY BOK
Autonomy and Morality: A Reply to Willaschek 204

DAVID WIGGINS
In a Subjectivist Framework, Categorical Requirements and Real Practical Reasons 212
DAVID GAUTHIER
Subjective Obligation: A Reply to Wiggins 233

ALLAN GIBBARD
Preferences and Preferability 239
JULIAN NIDA-RÜMELIN
Goodness and Rational Preferability: A Reply to Gibbard 260

Part III: Preference and Ethics

JOHN BROOME
Extended Preferences 271
RUDOLF SCHÜSSLER
Wish You Were Me: A Reply to Broome and a Comment on Harsanyi's Extended Preference Theory 288

RAINER HEGSELMANN
Experimental Ethics: A Computer Simulation of Classes, Cliques, and Solidarity 298
ULRICH KRAUSE
Solidarity among Rational Egoists: A Reply to Hegselmann 321

RAINER WERNER TRAPP
The Potentialities and Limits of a Rational Justification of Ethical Norms, or: What Precisely is Minimal Morality? 327
ANTHONY SIMON LADEN
A Hobbesian Choice: A Reply to Trapp 361

CHRISTOPH FEHIGE, RICHARD M. HARE, WOLFGANG LENZEN, JEFF MCMAHAN, PETER SINGER, THOMAS SPITZLEY, AND ULLA WESSELS
Symposium on Possible Preferences 367

- CHRISTOPH FEHIGE AND ULLA WESSELS
  Introduction to Possible Preferences 367

- PETER SINGER
  Possible Preferences 383

- RICHARD M. HARE
  Preferences of Possible People 399

- WOLFGANG LENZEN
  Who Counts? 406

- ULLA WESSELS
  Procreation 429

- JEFF MCMAHAN
  Preferences, Death, and the Ethics of Killing 471

- THOMAS SPITZLEY
  McMahan on Psychological Continuity and the Value of Future Goods 503

- CHRISTOPH FEHIGE
  A Pareto Principle for Possible People 508

Notes on Contributors 544

Name Index 550

Subject Index 555

CHRISTOPH FEHIGE

A Pareto Principle for Possible People

1. To Wish or not to Wish
2. Pareto-Superiority among Wishes (POPSAW)
3. To Be or not to Be
4. Antifrustrationist Utility (PAF)
5. The Format of a General Universal Pareto Principle (FGUPP)
6. The General Universal Pareto Principle (GUPP), the Mere Addition Paradox, and the Repugnant (and the Very Repugnant) Conclusion
7. GUPP: Selected Properties and Consequences
Acknowledgements and References

How good or bad is a world? Let us assume, as so often, that this is a matter solely of the preferences it contains and of their frustration and satisfaction.[1] One question we shall then have to face is how the existence of a preference and its satisfaction compares to the non-existence of this preference: is it better, or worse, or just as good, or sometimes one and sometimes the other?

Section 1 will argue at length that, ceteris paribus, the two options – satisfied preference and no preference – are equally good, a doctrine we can call antifrustrationism.[2] This settled, sections 2 to 7 will begin to translate antifrustrationism into moral principles, and to investigate the consequences.

Choices that make a difference to the number or identity of people who will ever exist – *different people choices*, as Derek Parfit termed them – will receive special attention.[3] Antifrustrationism helps us with these. For, within preference-based ethics, the value of people will depend on the value of preferences. Like any theory about the latter, antifrustrationism has significant implications for the former.

---

[1] If your favourite morality is less exclusive, and features preferences as just one component among others, you might read the paper as a reflection about that component. More on the relevant notion of preference, and on related concepts, in note 4 below.

[2] Brief references to antifrustrationist ideas have come up in the literature occasionally: Lockwood (1979), p. 164, McMahan (1998), sect. 6, Parfit (1984), sect. 124, Petersson (1993), sect. 2, Singer (1980), (1993), pp. 128 f., (1998), sects. 2 f. But the position has never been fully stated or defended, nor given a fair hearing. None of the foregoing authors endorse antifrustrationism or anything like it.

[3] The locus classicus for these issues is part IV of Parfit (1984); for the term "different people choices", see ibid., sect. 120.

It might be worth announcing that some of the principles we shall meet on the way, especially in section 5, should be of interest independently of anti-frustrationism; and that a more fine-grained guide to this paper can be found at the beginning of section 7.

1. To Wish or not to Wish

You ought to get, says preference-based ethics, what you want: it is good that, if individual *a* wants it to be the case that *p*, then *p*.[4] Since preferences can conflict, change over time, and vary in strength, we should be more precise: it is pro tanto good that,

> if, at point of time *t* and with strength *s*, individual *a* wants it to be the case that *p*, then *p*.

Let us call any conditional of this form a *Good Sentence*. Modifiers like "pro tanto", "ceteris paribus", or "intrinsically" should be read as saying that the truth of a Good Sentence is good in so far as *that* preference of *a*'s is concerned. We disregard other preferences, no matter whether the desideratum *p* affects, or even entails, the fate of some of them; by constituting other 'pro tantos' (being the topic of other Good Sentences, that is), their moral impact is sure to be registered separately. Similarly, the word "wants" refers to people's intrinsic wishes, not to what they want solely for the sake of something else. Most of the time, these are the sorts of readings we will have in mind for the moral and motivational words.

A Good Sentence being a conditional, there are two ways it could be true.[5] Either the antecedent and the consequent are both true – in other words, the preference exists and is satisfied; this we can call the ***orexigenic*** case, for it involves, somewhere along the line, the creation (γένεσις) of a wish (ὄρεξις). Or the antecedent is false – in other words, the preference doesn't come into existence in the first place; this we can call the *prophylactic* case. Are, ceteris paribus, both of these options, orexigenic and prophylactic, equally good?

---

[4] The usual stipulations apply: People need not be aware of their preferences; what counts is rather the attitude they ***would*** have towards something *if* they fully represented it. – People prefer to spend their conscious time pleasantly (cf. Kant 1785, pp. 415 f., Singer 1993, p. 131); more accurately: for any degree of pleasure they want to spend their conscious time pleasantly to at least that degree. This type of stipulation makes sure that preference-based ethics is sensitive to pleasure and the absence of pain. – To say that a preference is fulfilled or satisfied or frustrated is not to say that the preferrer's mind is affected; all it says is whether the desired state of affairs holds. – Words like "preference", "wish", "want", or "desire" usually refer to intrinsic wishes (a little more on this presently) and are used interchangeably.

[5] The appeal to the standard truth conditions for "if __ , then __" is a purely heuristic device and not essential to the argument; this is explained more fully below, at the end of this section.

Prophylacticism

We can make short work of *prophylacticism*: the claim that for at least some Good Sentences the prophylactic approach is, ceteris paribus, better than the orexigenic. Why should anybody think so? If the sun is shining anyway, then why shouldn't I want it to shine? How could, all by itself, the existence of a satisfied desire make the world a worse place? (Remember that any badness in the desideratum would *not* count as "all by itself".)

Sometimes, so the prophylacticist might reply, satisfaction is preceded by anxiety, despair, pain, restlessness, or thirst, or by a feeling of deprivation or emptiness. Sometimes satisfaction itself isn't satisfying at all, but disappointing. Sometimes it gives rise to other preferences that are hard, or even impossible, to satisfy. Sometimes others dislike the desideratum – one man's joy is the other man's sorrow. Sometimes what I want is scarce and my preference, being part of a demand that outruns the supply, creates a conflict of interests, maybe even a feud. Sometimes satisfying the preference requires disagreeable work from somebody. And so on.

But, firstly, preference-based ethics can recognize these drawbacks *without* being prophylacticist. It recognizes them anyway, since they all imply the frustration of various *other* preferences: for feeling good, for buying the only Picasso on sale, against dirty work, etc. In fact, that is what makes them drawbacks.

Secondly, the inconveniences, frequent as they may be, cannot support prophylacticism as long as, *conceptually* speaking, they are incidental to the preferring and the satisfaction – in which case, far from establishing ceteris paribus badness, they just violate the ceteris paribus condition. Even if causally linked to preferences, the drawbacks could not argue against preferences as such. ("Intrinsically good" is immune to causal reverberations just the way "intrinsically pleasant" is: even if sex caused cancer, it would not follow that sex as such isn't fun.)

Prophylacticism would have to view at least certain preferences themselves, regardless of their side-effects, as *diseases*, and their satisfaction as the cures. The person who is spared from the disease (read "preference") is better off than the person who had the disease, but was cured (read "got what she wanted"). This sounds plausible for diseases, most of which are unpleasant by definition. But there is nothing about satisfied preferring *per se* that would bear out the analogy. Preference *plus, say, worry* plus satisfaction might bear it out, but just preference plus satisfaction does not.[6]

By itself, the prophylactic truth of a Good Sentence is not better than the orexigenic truth. What remains to be examined, then, is whether it is worse; or equally good; or whether this varies from one Good Sentence to another. To find out, we distinguish two cases: either the designated bearer of the new preference is a preferrer anyway (case I) or she is not (case II).

Orexigenesis vs. Prophylaxis I: The Case of the Preferrer

What if individual *a*, the Good Sentence's protagonist, is a preferrer anyway? In this case (which will now occupy us for a few pages), we can consult the spirit of preference-based ethics, the very idea we set out to capture by talking of "preference satisfaction": that the good is made up from what is good *for* individuals. So we have to check whether it is good *for the preferrer* to get a new and satisfied preference.

(Why just for the preferrer? What if somebody else, say *b*, wants *a* to have more preferences? Wouldn't that make them a good thing? Yes, extrinsically. We ought to fulfil *b*'s preference, and in the case at hand *a*'s extra preferences would be a means to that end. But are they an end in themselves? Since that is our question, we may leave *a*'s fellow-preferrers out of the picture.)

Now, if somebody induces in me the wish to own a Ferrari, and makes sure I get a Ferrari, am I ipso facto any better off than I would have been without the wish?[7] It is hard to see why. Certainly, I could sell the Ferrari and spend the money on other projects that I had anyway, but that doesn't count as "ipso facto"; the Ferrari would then have been instrumental in making another Good Sentence true of me – "if Fehige wants to own a house, he owns a house", or whatever. But we are still after the intrinsic, non-instrumental value of the new and satisfied preference.

What if the Ferrari preference gives me an aim in life? I now know what I'm working for, and have something to look forward to. Every time I put a thousand dollars in my Ferrari savings account I have some sense of achievement, and I have a huge lump of that sense the day I can pick up the car from the dealer. It has been, all in all, good fun to work for the result, and it is good fun to experience it. The new preference has cheered me up – isn't that good?

It is. But this story, too, just shows that a satisfied extra preference can be a means to a good end. Roughly speaking, hedonic happiness is a state of mind that people want to experience: it is a conceptual truth that any preferrer wants, ceteris paribus, to spend her conscious time as pleasantly as possible.[8] So it is once again another preference, the one for feeling good, that the Ferrari preference has helped to satisfy; once again we have failed to detect any independent virtue in its existence.

---

[6] Cf. Sidgwick (1907), book I, ch. IV (§ 2 and end-note), and book IV, ch. I, § 2, fn.

[7] Parfit (1984, appendix I) employs a similar question to a different purpose, cf. sect. 7 below.

Now suppose that I would *start* loving the Ferrari if I owned it; that is, the wish to receive the car at point of time t was not there before, but would be there afterwards. Would not the existence and satisfaction of the additional ex-post preference be a good thing?[9] But we must remember that all other things are equal; so we have to compare the scenario featuring both the Ferrari and the ex-post preference to a second scenario *in which I feel precisely as good as in the first one and in which I never want or miss either the Ferrari or anything it would have procured me.* If this is so, why would we want to say that I'm worse off in the second scenario? And why should, of all satisfied preferences, ex-post ones have intrinsic value, whereas others don't? It would be curious to claim that the combination of preference and satisfaction is a good thing if and only if the second component is delivered before the first.

All this is not to say that nobody wants new and satisfied preferences, or that it would be irrational, or morally irrelevant, to want them. Mary's *wish* to have additional wishes conveys instrumental value on their existence – the value of avoiding the frustration of *another* (viz. the higher-order) wish. *We respect and count every orexigenic metapreference.* The argument here is not a crusade against wanting to want more. All it rebuffs is the claim that, even if people *don't* want to want more, wanting more (and getting it) would *benefit* them. Whenever we talk of the ceteris paribus value of new and satisfied preferences, we must remember that a metapreference for or against them is a ceterum impar; we must, that is to say, remember to pretend that neither their existence nor their non-existence are desired. For, if either of the two were desired, it would *thereby* be of value.

Has the choice of the example influenced our judgement unduly? The more bizarre the wish, the more tempting the verdict of futility. Are satisfied extra preferences worthless *only* if they are for silly things like Ferraris? One might be inclined to say so; parents, for instance, are glad if their child develops certain preferences instead of others, or instead of none, and it is true of some motivations, as opposed to others, that a lack of them would be regarded as disadvantageous, evil, or pathological. But why? Why would it be a bad thing that, say, your child will never want to go to school, or will never want to work? Because in that case neither the child nor you nor the taxpayer are likely to be happy. In the world as it is, the absence of wishes for literacy and for work tends to increase preference frustration. However, if this were not so – if all other things were indeed equal –, then it is hard to see how even these extra wishes could make somebody better off. How could they fail to be in the same class as those for Ferraris?

---

[8] A little more on this in note 4 above.

[9] A variation of this question would adduce not the ex-post preference for the car, but for having the satisfied preference to own the car. The reply would be essentially the same.

Content-based discrimination is improbable as a component of either 'good for' or 'good' anyway. Preference-based ethics assumes that these concepts are concerned with a *subjective* magnitude. For them, it's the preferredness that matters. And if it is only preferredness that conveys value on objects in the first place, then it is unlikely to matter whether these are books, Ferraris, games of pushpin, or poems.[10] Look at any preferrer and leave aside specific metapreferences he might happen to have already – how can we deny that every set of preferences that would yield the same pattern of frustration and satisfaction would be equally good *for him*? And if so, that, ceteris paribus, they are equally good simpliciter? Since ethics turned to preferences precisely in order to *avoid* discrimination, it won't do to now discriminate *among* preferences.

In particular, we must be on our guard against the grip that our purely personal metapreferences may have on our moral judgements. *We* are people who want to have certain preferences (e.g., for reading), and *we* would *therefore* be better off having them. Judging from this that these or other extra wishes would be good for people who *don't* want them is just as sound as an inference from *our* desire for, say, rhubarb to the unconditional conclusion that everybody – no matter whether they want it – ought to eat, or would be better off eating, rhubarb.

The claim that, pro tanto, satisfied extra preferences don't benefit people resembles what the sages have always told us. We should, so their advice goes, want less than we usually do. Why? Parts of the Stoic arguments were empirical and needn't concern us here. (In the world as it is, lots of desires are painful, lots of satisfactions short-lived and hard to come by; etc.) Another part, however, was this: even if preferring didn't hurt, and satisfaction were guaranteed, what would be the point? What good is the pair of them? At best, it will restore my welfare to what it had been without the preference. One more time: suppose we paint the tree nearest to Sydney Opera House red and give

---

[10] Cf. Bentham (1825), p. 253; today this is known as the "condition of neutrality", see, for instance, Kern/Nida-Rümelin (1994), sect 6.1. Criticizing a draft of this paper, Klemens Kappel (1996, sect. 3) says that some satisfied extra preferences benefit the preferrer and some don't. But he gives us no idea which preferences are supposed to fall in one class, which in the other, nor why.

Kate a pill that makes her wish that the tree nearest to Sydney Opera House were red – have we really done her a favour?[11]

A similar argument could be made from rational decision making. If, in themselves, satisfied extra wishes were good for preferrers, then prudence, too, since it is in charge of its clients' welfare, would have to be orexigenic. Rational decision theory would have to be supplemented with something like

The Orexigenic Axiom:

> Other things being equal, a rational preferrer wants to acquire new and satisfied preferences.[12]

But should we let decision theorists get away with this? Probably not. We would say that, yes, a rational person *can* have these orexigenic higher-order desires, but she need not. The matter is not settled by rationality itself, but is part of the arational input.

The view that satisfied extra preferences benefit people would have bizarre implications in the field of intrapersonal *aggregation*, too. Suppose, for example, that John, who loves life and has lots of projects under way, goes to the doctor and is told, out of the blue, that he is suffering from a rare virus disease and will die in three months' time. "That, however, is very good news; for", says the doctor, a keen believer in the value of satisfied extra preferences, and a reliable person, "the virus you've caught has another effect as well. Next week or so, you will begin to develop an immense desire that after your death ten tons of pink foam rubber be deposited in your front garden. I will see to that right after your burial, and your health insurance will pay for it. Of course, your death will frustrate some other wishes of yours, including past ones, and some of these, like your wish to survive, are rather strong; death will also deprive you of the value of all the satisfied extra preferences you would have developed had you not caught the fatal virus. But I assure you that all those wishes together are not as strong as the satisfied extra one that is in store for you, the one for pink foam rubber. It will make up for all of them. It's a pity most people are immune to this virus, but we are working hard to make it more widely available." The doctor congratulates John. But I don't agree. John, I think, has little reason to rejoice.

It could be objected that John's new desire is not a "global" one – none, that is, "about some part of one's life as a whole, or [...] about one's whole life"[13] –, and that therefore it cannot outweigh the global desires frustrated by his death. Or that preferences for what happens after one's death do not affect one's welfare anyway. The upshot of these objections would be that, if we give priority to global preferences, or disregard preferences concerning the time after the preferrer's death, we can favour orexigenesis without having to say that John the moribund is a lucky man. It could also be objected that the person with an extra wish as bizarre as that for pink foam rubber is ipso facto (or at least to that extent) no longer John, and that we're therefore not talking about an increase of *John's* benefit.[14] But even if these moves were plausible, think how the example could be modified to avoid them: give thirty-year-old John a few more *years* instead of a few months, and let him develop the wish to spend the autumn of his short life as owner of some specific item which is not totally at odds with his character; if, for instance, John has always loved motorbikes the new desideratum could be our old friend, the Ferrari sports car. The extra desire is then global, concerns the preferrer's own life-time, and leaves his identity unscathed. Do we share the doctor's outlook *now*?[15]

---

[11] If we had, then Christmas could look different.

[12] If you think decision theory is concerned with the rationality of actions rather than of preferences, then you can reword the Orexigenic Axiom and my comments accordingly. The Orexigenic Axiom is also discussed in Wessels (1998), sect. 3.3.

Ceteris paribus, we have no moral reason to subject preferrers to orexigenesis rather than to prophylaxis. Over the last few pages, this conclusion has begun to look safe. Before closing the file on preferrers, however, we should briefly look at them when they are unconscious, a case already halfway to that of the non-preferrer. Suppose, then, that a preferrer is unconscious (asleep, say, or in a coma) and that it is up to us whether she will die as she is, or wake up to a normal life. To real-life cases of this type, moral disputes about the value of extra preferences will hardly make the difference. There will always be plenty of non-extra preferences, hers or other people's, to generate obligations to get her back on her feet, preferences that have been, are, or will be around anyway. The woman on the Clapham omnibus has, before she falls asleep or out of the bus, preferences (at least implicit ones) to wake up, or, at any rate, preferences that, in the world as it is, could be satisfied if and only if she woke up. She is in the midst of all sorts of things she wants to get done: buying a house, bringing up her children, making provisions for a happy future. There will also be friends and relatives who want her to wake up, and would feel good if she did and terrible if she didn't. For such reasons, even people who don't believe in the value of extra preferences will think that, on *our* planet, it is wrong to kill unconscious preferrers if we can give them a happy life instead.

---

[13] Parfit (1984), appendix I; see also sect. 7 below.

[14] Cf., in a slightly different context, the thoughts about identity, preferences, and welfare in Bricker (1980), towards the end of sect. IV.

[15] Don't forget to abstract from the fun the Ferrari might be. The desire to have fun while one is conscious is not an extra one – it is there anyway; cf. our earlier Ferrari discussion.

This said, we can proceed to a coma patient, or to a sleeper, in never-never land, where everything is different.[16] Here she is aimless and has nothing to live for. Never-never land is *designed*, by moral philosophers, to make sure that not a single non-extra preference argues against the person's death. It is designed to turn the question of orexigenesis into the only and therefore decisive question: if anything at all, then just the value of the extra preferences she *would* develop if we woke her up could tip the moral scales in favour of her survival. Exotic as this may be, in the never-never land death scenario *neither the preferrer nor anybody else ever had, has, or will have any preference, not even an implicit one, for anything that requires her survival.* This implies that, if we could and did ask her now (and had asked her before), and if she fully represented the option now (and had fully represented it before), her own answer would be now (and would have been before) that she's not interested in a happy survival. But then the claim that death would *harm* her sounds wildly implausible. Once again, orexigenesis turns out to be, all by itself, not better than prophylaxis.

Orexigenesis vs. Prophylaxis II: The Case of the Orectic Novice

So much for the case of the preferrer. Next, suppose that *a*, the hero of our Good Sentence, is an orectic novice – a young fetus, for example, or a yet unconceived individual. It is up to us whether individual *a* will ever have *any* preferences, even implicit ones. Here the question whether we ought to make an *a*-instance, or several *a*-instances, of the Good Sentence true *by orexigenesis* amounts to the question whether we ought to put *a* on the orectic map *at all*.

We have reason to believe that there is no such obligation.[17] Firstly, remember from the previous case that morality doesn't prescribe orexigenesis for preferrers. So if we now prescribed it for non-preferrers, we would prescribe an initial set of satisfied preferences, and that initial set only. But this would be a puzzling asymmetry. Imagine a theory saying that everybody ought to keep their first promise but needn't bother about the further ones. No matter what you think about promises, the idea that only the first one counts is peculiar. Ditto for satisfied preferences.

---

[16] This case is also discussed in Wessels (1998), sects. 2.2 f.

[17] Some of the best-known reasons are purely intuitive, and I do not wish to count on them. One of them is that, if we accept this sort of obligation, we shall have a hard time avoiding the so-called Repugnant Conclusion, to be met in sect. 6 below. Another argument says that, if morality favoured orexigenesis, most real-life women would be under an obligation to procreate (and, if pregnant, not to abort); and that, since this would be counter-intuitive, morality should not favour orexigenesis.

Secondly, morality strives to go by the affected parties' preferences. If the affected party is a non-preferrer, there *are* no preferences to go by. What we need, then, is to appoint some sort of guardian who will tell us what is best for her ward, the non-preferrer. But whom? The obvious candidate is rational decision theory; usually, it has some input to process, but we can treat input that isn't there as indifference, and see what the theory comes up with. Since we know (from the discussion of the preferrer's case) that it has no orexigenic axiom, it will – in the absence of arational orexigenic *input* – give no positive weight to satisfied extra preferences. Thus, the guardian will let morality know that satisfied extra preferences, and hence birth, are of no use to the potential preferrer.

(This sort of argument, it might be objected, would establish that we may do anything we like with the non-preferrer. Can we drag him even into a miserable existence just because the ex-ante absence of a preference against misery would be counted as indifference towards misery? No, and thus the objection fails. Frustration is something that pure rationality advises us to steer clear of: ceteris paribus, the rational preferrer wants not to acquire preferences that will be frustrated. For rational decision theory to rule out misery, then, it takes no input. The guardian does not lack benevolence; all she denies is that benevolence is orexigenic.)

Thirdly, alternative views – views according to which we ought to bring it about that individuals have preferences – have little to recommend them.[18] They come in two flavours. One is that we ought to *create* individuals with satisfied preferences. Support for this claim has been seen in the Bible (*Genesis* 1:28, 1 *Timothy* 2:15, etc.), in some people's intuitions (opposed, as so often, by other people's), and in precisely one argument, devised by Richard Hare.[19] Hare argues that to be moral is to have analogous preferences for analogous situations; that some real-life people want to have been born; and that, hence, morality requires them to have, for analogous situations (in which *other* people's birth is at issue), analogous preferences (preferences for *those* people's birth, that is).

The second type of alternative doesn't go quite as far. It says that, *once there is* an individual, we ought to make sure it has satisfied preferences.[20] Like the previous claim, it entails that abortion is wrong (if it excludes the fetus from a life of more satisfaction than frustration); unlike the previous claim, it does not entail duties to procreate. In her contribution to this symposium, Ulla Wessels has shown the two positions to be either untenable, or unjustified, or both.

---

[18] More about them in Fehige/Wessels (1998).

[19] See Hare (1975) and (1998), as well as essays 5 f. and 11 f. in Hare (1993). The claim, though perhaps not Hare's argument, is supported, mutatis mutandis, by Bill Anglin, John Leslie, Yew-Kwang Ng, T. G. Roupas, Richard I. Sikora, J. J. C. Smart, and others.

[20] Variants of this view are held, with important qualifications, by Wolfgang Lenzen, Jeff McMahan, Don Marquis, Mary Anne Warren, and others; see, for instance, Lenzen (1998) and McMahan (1998).

The three arguments settle our second case, that of the non-preferrer: we have obligations to make preferrers satisfied, but no obligations to make satisfied preferrers.[21]

Antifrustrationism

Our discussion of orexigenesis and prophylaxis in general is thus concluded. The upshot is that, intrinsically, the two of them are always equally good. We don't do any good by creating satisfied extra preferences. What matters about preferences is not that they have a satisfied existence, but that they don't have a frustrated existence. Let us call this result *antifrustrationism.*

Today's point of departure was the fundamental creed of preference-based ethics: it is pro tanto good that, if somebody wants something, he gets it. It follows from antifrustrationism that the creed should indeed say no more than what, according to classical logic, it does say; viz., no more than that it is good that it is not the case that, though *a* wants that *p*, *p* does not hold. Less clumsily: it is good that *a* has no frustrated preference for *p*.

Maximizers of preference satisfaction should instead call themselves *minimizers of preference frustration.* Analogously for non-maximizers, i.e. non-utilitarians. The search for the best social welfare function of any type, utilitarian or not, must be the search for the best distribution of preference frustration, or rather of *avoided preference frustration.* We can go on saying that what's at issue is the best distribution of *utility* or *welfare*; but then we shouldn't just say that utility or welfare is a matter of *preference satisfaction.* This statement is open to many interpretations, including prophylactic, orexigenic, antifrustrationist, and various mixed ones. Again, the adequate way to disambiguate is to say that by "utility" or "welfare" we mean a measure of avoided preference frustration.

Two misunderstandings are worth warding off. Firstly, antifrustrationism should not be confused[22] with the grotesque view that a person's *future* preferences do not count. Of course they do – just not orexigenically. I have ceteris paribus no obligation to bring it about now that you will have preferences later. But I do have obligations to do now whatever helps to satisfy the wishes you will have later; this includes the obligation, where it is up to me now *which* preferences you will have later, to give you the set of preferences that, reckoned throughout your life-time, will result in the least frustration for you.

---

[21] The view and this type of slogan originate with Jan Narveson, see his (1967a), pp. 47–50, and (1967b); Vetter (1969) and (1971) agrees; Sprigge (1968) disagrees, as do of course the authors mentioned in note 19 above. The position presented here generalizes and, in a sense, justifies Narveson's claim by transferring it from the level of people to that of preferences.

[22] As it seems to be by Lenzen (1998, sect. 5.2).

Secondly, note that the argument for antifrustrationism does not depend on the 'paradox of material implication'. The somewhat disputed convention of calling a conditional true if its antecedent is false was convenient in that it enabled us to *present* the inquiry as being into the ways Good Sentences can be true. But the *moral* comparison was that between the various ways truth and falsity could be spread over a Good Sentence's components, regardless of how they affect the truth-value of the overall conditional. So if you are suspicious of the classical "if, then" (or of its role in monadic deontic logic), never mind – the question had nothing to do with it. It was whether preference plus satisfaction is better than no preference. And the answer is no.

2. *Pareto-Superiority among Wishes (POPSAW)*

The remainder of this paper will present a few uncontroversial elements of antifrustrationist morality; they are uncontroversial in the sense that somebody who accepts the conclusion from section 1 is unlikely to reject them. (Note that some of these elements ought to recommend themselves to other moralities as well, see especially sect. 5.) To begin with, there is

The Principle of Pareto-Superiority among Wishes (POPSAW):

(i) If the Good Sentences true in world α form a proper subset of those true in world β, then β is better than α.
(ii) If the Good Sentences true in world α form a subset of those true in world β, then β is at least as good as α.

Some explanations. For preference-based ethics, Good Sentences are the atoms of the good. Every such atom is good, and all that is good is, in the end, made of such atoms only. Now, if a world has all the atoms of the good that another world has, *and some more*, then how can it fail to be better? It's like a heap of gold nuggets. If you simply *add* gold nuggets to an existing heap, how can the resulting pile fail to be more valuable?

In other words, if we think of preferences as little creatures, shouting "I want to be satisfied!", then what POPSAW says is this: "If none of these creatures would veto the choice of world α over world β – i.e. if none of them would say, 'Hey, I'm against it! I'd be satisfied in α, but frustrated in β!' –, then β can't be worse than α." This little dramatization explains the "Pareto-Superiority" in POPSAW's name. We remember that a world is called Pareto-superior to another one if and only if it is at least as good as the other one for everybody, and better for somebody. Pareto-superiority is generally held to be a sufficient criterion for betterness;[23] claims of this type are known as Pareto principles and will appear again in section 5.

Notice how weak POPSAW is. To a large extent, ethics is about *conflicts* of interest: what ought to happen when preferences clash, when we have to choose, that is, between sets of Good Sentences where *neither is a subset of the other*? POPSAW does not say anything about choices of this type. It does not *weigh* wishes against each other.

In particular, POPSAW should not be confused with a principle saying that the *more* Good Sentences are true in a world the better. Firstly, it would have to be explained how one is to count preferences. Secondly, such a principle would be about as controversial as utilitarianism is, and for roughly the same reasons. It would, other things being equal, rather dump the frustration of 1000 preferences on one person's back than inflict 334 frustrations on each of three people. But this is a type of moral judgement as debatable as they come.

POPSAW refuses to comment on in*tra*personal problems of preference aggregation in the same sense in which we saw it refuse to comment on the in*ter*personal ones. Asked, for instance, to choose between frustrating either an earlier or a later preference *of Mary's*, POPSAW has, because of the time parameter in the Good Sentences, nothing to say. Even there it is neutral.

Another part of the Good Sentences, the strength parameter, drops into the right place as well. If your concept of preference doesn't allow for a notion of strength, then never mind – you can simply delete strength from the Good Sentences. Suppose, however, that you do allow for a notion of strength. You might then want to say that the strength of a preference makes a difference for the goodness or badness of its frustration or satisfaction. You are, for example, likely to believe that, ceteris paribus, frustrating a stronger preference is worse than frustrating a weaker one. If you want to say any such thing, then Good Sentences without a strength parameter would not be sufficiently fine-grained to assure POPSAW's adequacy. To see this, look (for any contingent proposition p, any point of time t, and any preferrer a) at two non-p-worlds α and β that differ in no other orectic respect than this: in world α, individual a has, at t, a weak preference that p (say of strength 1, whatever you mean by that), and that preference is frustrated; in world β, *a* has, at *t*, a strong preference that *p* (say of strength 10), also frustrated. Now, if Good Sentences were strengthless, they would not register the difference; the ones true in world α would be a subset of those true in world β, and vice versa, and hence clause (ii) of POPSAW would declare α to be at least as good as β, β to be at least as good as α, and the two, therefore, to be equally good – whereas you think they aren't, because frustration is stronger in β than in α. *With* the strength-parameter, POPSAW avoids the moral judgement you disagree with. The Good Sentence "(*a* wants at *t* with strength 10 that *p*) → *p*" is true in world α (since the antecedent is false there), but false in world β; vice versa for "(*a* wants at *t* with strength 1 that *p*) → *p*". So there is no subset relation either way; POPSAW, therefore, says nothing, and hence nothing inadequate, on the comparative quality of α and β.

---

[23] Cf. Pareto (1906), ch. VI, sect. 33, Sen (1970), sect. 2*1.

POPSAW also implies that, intrinsically, the orexigenic and the prophylactic way of making a Good Sentence true are equally good. The different *ways* of truth are just not mentioned in the principle and thus can't leave moral traces in it. More precisely: consider any two worlds with the same set of Good Sentences true in them, but perhaps true in radically different ways, as far as orexigenesis and prophylaxis are concerned. Then clause (ii) of POPSAW will say that the first world is at least as good as the second, and the second at least as good as the first; hence, that they are equally good. As was shown in section 1, this feature, viz. the moral equivalence of prophylaxis and orexigenesis, is welcome – it's antifrustrationism.

So much for Pareto-superiority among wishes. If all the wishes frustrated in world α are frustrated in world β, and some more, then β is worse than α. Wishes were the gate through which we made Pareto enter the realm of the possible. In that realm, he will have to make his way from preferences to preferrers, a journey that will begin in section 4. But first, some implications of POPSAW call for our attention.

3. *To Be or not to Be*

POPSAW is not very enthusiastic about existence (but then, who is?); the principle entails, for example:

(i) Nothing can be better than an empty world (a world without preferences, that is).
(ii) Our world is worse than an empty world.
(iii) It is ceteris paribus wrong to create a being that will have at least one unfulfilled preference.

This trio is likely to violate at least some people's intuitions,[24] and deserves consideration. I will start with a few general remarks and then look at each of the claims in turn.

The first of the generalities is that POPSAW stands on firm ground, as I hope to have shown at some length in the previous two sections; whoever dislikes any of its implications should tell us where these arguments went wrong.

Secondly, it is doubtful anyway whether we should go by what we intuit. Thirdly, even if we do, a principle should not be disqualified for hurting *some* intuitions; it might still violate *fewer* than others and thus win the competition. This is especially true for population ethics, where theories that obey everybody's intuitions just don't exist – the only thing that several decades of research in this field can safely be said to have established. Fourthly, most people's intuitions with respect to different people choices are even inconsistent (see sect. 6 below); it follows that *any* consistent theory will have to make most of us say a few things that, prima facie, we won't be inclined to say. Fifthly, I for my part find none of the claims (i) to (iii) counter-intuitive, and I'm in good company. The Buddha, for instance, seems to be on our side. POPSAW explains the widespread belief that existence is a game we cannot win, and are most likely to lose.

With these generalities in place, I return to the individual claims (i), (ii), and (iii). As to claim (i): Remember that nothing that has been said so far excludes the possibility of non-empty worlds that would be, according to POPSAW, of the same value as empty worlds; they would be non-empty worlds in which every preference is satisfied. Note also that the only alternative to claim (i) is obligations to procreate – and now who's being counter-intuitive?

As to claim (ii): Firstly, it does not imply that we should empty the world we live in; we can't, for there have been preferences already. (An empty world is one that features no preference throughout its history.) Secondly, suppose your moral sentiments do indeed find fault with claim (ii). It would then be interesting to know how many happy lives on earth you think *have* sufficed to outweigh Auschwitz. How obvious is it that the whole thing – famines, world wars, and gas chambers included – is 'better than nothing'?

As to claim (iii): Note, as always, the "ceteris paribus" clause. Whenever there are people who want to have a child, the cetera are no longer paria. POPSAW does not prescribe childlessness to would-be parents.

Especially when confronted with claim (iii), our intuitions must be on the look-out against two misleading thoughts. One of them could be expressed like this: given a happy person with lots and lots of satisfied preferences, how could *one* frustrated preference suffice to make the difference? The answer is that it would be surprising indeed if one tiny frustration could turn a decidedly good thing into a bad thing. It can't. Instead, the existence of the person with all her satisfied preferences is morally *neutral*, and one tiny preference frustration is one tiny departure from neutrality towards badness. Think of a barrel that is filled with water right to the rim. Can removing just one glass of water make the barrel less full than it could be? The answer is yes.

---

[24] Some of these issues are mentioned or discussed in the works listed in notes 2 and 21.

It is true that we are used to thinking of existence as a big deal – just look at birthday parties. But if antifrustrationism is right, then, in the thought experiment that features an entity with just one frustration, existence is, other than in real life, no big deal. In the thought experiment, it makes no major moral difference. You can't do any better than having no frustration, and the person who will have just one does only a *little* worse, just like the barrel of water is only a *little* emptier if you take out just one glass.

Note also that, if there were a disproportion between the trifle (just one frustration) and the matter it settles (is the life in question worth while?), then practically every morality would exhibit the same sort of disproportion. Antifrustrationist or not, ethics is bound to say that there is *some* level of well-being (or rather, of unwell-being) below which existence is a bad thing. No matter where that threshold is – if we imagine a person on the brink of it, we shall have to say that one more frustration would be the straw that breaks the camel's back.

The other misleading thought would be that of POPSAW as a sword of Damocles over a happy person's (or population's) head: one frustrated wish, and POPSAW will chop off her head. Unpleasant ideas of losses and victims creep up in us, as if POPSAW wanted to *put an end* to an enterprise that, now it has started, is perhaps good fun, or at least tolerably so. But POPSAW plans nothing of the sort. It doesn't say that you should be killed because you have toothache (or cancer, or whatever), or that mankind should stop procreating because future people will have toothache (or cancer, or whatever). A single preference for survival or procreation, and our principle doesn't interfere. POPSAW permits the show to go on as long as there are, or if there ever have been (as indeed there have), people who want it to go on.

4. *Antifrustrationist Utility (PAF)*

Most of the time, preference-based ethics is conducted not – as it has been in this paper so far – in terms of an individual's various wishes and their fulfilment, but in terms of a magnitude made up from them: the individual's utility, well-being, or some such thing. It's all very well knowing about when you desired what and whether you got it, but at some point in our moral thinking we will want to sum up all this information and say how, all things told, you are doing. We will want to work with utility functions: functions that assign to an individual *a* and a possible world α a value $u_a(\alpha)$, where "*u*" is mnemonic for "utility", that represents how well off *a* is in world α. Thus, antifrustrationists will have to ask themselves what it is for a concept of utility to obey, as we argued it should, antifrustrationism. At least a part of the answer is that it should satisfy

The Principle of Antifrustrationism (PAF):

(i) If the Good Sentences true of individual *a* form a proper subset of those true of individual *b*, then *b* is better off than *a*.
(ii) If the Good Sentences true of individual *a* form a subset of those true of individual *b*, then *b* is at least as well off as *a*.

PAF is just POPSAW, from section 2, sliced into people. POPSAW looked at the Good Sentences true of all possible people and ranked entire worlds; PAF does the same thing with just two people and ranks just their well-being.

We can briefly review the old messages in the new format. First, let us pretend that PAF talks about preferrers, and about preferrers only. For that case, our very argument for *moral* antifrustrationism had proceeded precisely *via* the sort of individual antifrustrationism that PAF now asserts. Remember the Ferrari from section 1. We argued that satisfied extra preferences have no intrinsic value *for the preferrer*, and that orexigenesis and prophylaxis are of equal value *to the preferrer*. And that is what PAF says.

So much for PAF and preferrers. Now, what if we allow PAF to talk not only about a preferrer's welfare but about that of *any possible object*? How well off is a possible individual *a* with a world α if she does not exist in α or is not a preferrer in α? The question sounds odd. But if we don't take it too literally, it is quite respectable: can we *assign* utility levels to non-preferrers such that moral principles that employ utility levels will generate adequate results, and nothing but adequate results?

We can, and PAF does. PAF makes sure that *everybody whose preferences are all satisfied* is assigned a maximum level of utility. The expression "everybody whose preferences are all satisfied" covers various cases: the individual exists and has preferences and they are all satisfied, or it exists but has no preferences (say, it's a stone, or a chair), or it does not exist.

The inclusion of the last two cases, viz. of the non-preferrers, is not a trick. The justification and the implications of this convention are all among those of antifrustrationism and of POPSAW, and were presented and discussed in sections 1 to 3. We are merely translating the earlier thoughts into utility jargon.

In fact, the convention has little to do with full-fledged antifrustrationism. Suppose you had never heard of this doctrine, only of Narveson's slogan that we have obligations to make people happy, but no obligations to make happy people.[25] If you tried to transcribe this slogan alone into utilese, you would end up exactly with PAF's suggestion to assign maximum utility to all merely possible people: saying, with Narveson, that you can't do any good by turning a non-preferrer into a preferrer already *amounts* to saying, as PAF does, that nobody can be better off than a non-preferrer. (Since, if somebody could be better off, you *would* be doing a good thing by turning the non-preferrer into such a somebody.)

Summing up, then. PAF, the Principle of Antifrustrationism, is an elementary antifrustrationist constraint on the concept of utility. And it permits us – this is just one of its many blessings – to stipulate that "$u_a(\alpha)$", i.e. "*a*'s welfare in world α", is defined for *every* possible individual *a* and *every* possible world α – *even if* a *is not around in* α, *or is not a preferrer in* α. For, even if she isn't, PAF makes the right thing (the antifrustrationist thing) happen: she's assigned a utility that cannot be surpassed by anybody's utility in any world.

5. *The Format of a General Universal Pareto Principle (FGUPP)*

Let us call a doctrine *possibilistic* if it invites us to interpret it as assigning a utility level to non-existence (or – in the metaphysically more risky jargon we have used so far – to merely possible preferrers). Antifrustrationism is, as we have just seen, possibilistic. But so are other doctrines. That of Richard Hare, for example, can be expressed as assigning utility zero to merely possible preferrers, as well as to preferrers whose lives contain 'as many' satisfied preferences as frustrated ones; positive utility to all lives that contain 'more' satisfied than frustrated preferences; and negative utility to the others.[26] That PAFians aren't the only possibilists is important. Firstly, it might add to the acceptability of possibilism (and hence, to some extent, of PAF itself). Secondly, the present section should be of interest to *every* possibilistic doctrine, antifrustrationist or not.

---

[25] Cf. note 21.

[26] "as many" and "more" in scare quotes, since strength will have to be taken into account as well as number. Hare's theory is set out in the writings listed in note 19 above. The closest welfarist counterpart to the doctrine sketched out by Gregory Kavka in his (1982) is another example of possibilism. It can be expressed as differing from Hare's conventions only in that it assigns some positive value *k* to mere possibilia.

If every possible individual is assigned a utility in every world (no matter whether she's in it), then this is a real treat and we should stop for a moment to savour it. In a sense, it enables us *to look at all choices as same people choices*. We will be able to pretend that, in a sense, the set of parties that have moral standing is the same in all possible worlds: it's simply the set of all possible individuals. Putting things this way may be somewhat unusual, but is sound and convenient.

To explore some of the treasures thus opened up, let us start with the run-of-the-mill Pareto principle (mentioned towards the beginning of sect. 2) and see how far possibilism, together with an idea from Suppes (1966), permits us to strengthen it.

Pareto as We Know Him

The Pareto principle as we know it only compares worlds if their populations are identical. (By a world's "population" or "inhabitants" we will continue to mean its *actual* inhabitants; similarly, and notwithstanding the *façon de parler* from two paragraphs ago, the term "same people choices" remains reserved for choices with the same set of *actual* inhabitants in all the outcomes.) For worlds α and β that have the same population, say Ω, here is

The Ordinary Pareto Principle (OPP):

(i) $\forall x \in \Omega\ u_x(\alpha) \geq u_x(\beta)\ \ \&\ \ \exists x \in \Omega\ u_x(\alpha) > u_x(\beta)$
$\Rightarrow \alpha >_{\heartsuit} \beta$

(ii) $\forall x \in \Omega\ u_x(\alpha) \geq u_x(\beta)$
$\Rightarrow \alpha \geq_{\heartsuit} \beta$

[In prose:

(i) If all the parties are at least as well off with world α as with world β, and at least one of them is better off with α, then world α is better than world β.

(ii) If everybody is at least as well off with α as with β, then α is as least as good as β.]

Possibilizing Pareto

The first thing possibilism allows us to do is to take OPP as it is but say that Ω is the set of all possible individuals, and thus that the population of worlds

α and β need not be identical in order for Pareto to be able to compare them. Call this the *Format of a General Pareto Principle, FGPP.* "General" since it liberates Pareto from the realm of same people choices; "Format" since its moral *substance* will vary considerably with the choice of a possibilistic doctrine – the choice, that is, of the utility level to be assigned to non-preferrers.

Identity Bijected Away

FGPP, coupled with any version of possibilism, is already vastly more applicable than OPP. Next, let us get identity out of the way. What we want to achieve is universalizability (also known as anonymity, equality, impersonality, or symmetry): the ideal, widely accepted in ethics, that it must not matter who plays which part. Here's a trivial example. A world that features no sentient life other than John is as good as a world that features no sentient life other than Jane, as long as Jane's welfare in her world equals John's welfare in his. This is the sort of sense in which justice is blind.

A more interesting example is best conveyed by drawings of a type familiar from Derek Parfit's writings. Let each of the two diagrams in figure 1 represent the welfare profile of a possible world. (Ignore the arrows for the moment.) Every inhabitant of the world is represented by a vertical dash, and the length of the dash shows the person's welfare in that world. If a dash is above the horizontal line, then the corresponding person herself considers her life worth living; if not, then not. World α, for instance, as shown in the left hand diagram, has two inhabitants, a_1 and a_2; they are both far above the 'zero line', and a_2 is better off than a_1. When we compare worlds α and β, the Pareto in us, or at least a close relative of his, says that clearly α is better than β. The reason is universalist. It is not that any particular person (or welfare dash) in α stands in any particular relation to a particular person (or dash) in β. (a_1's dash, for example, is not larger than b_1's.) The reason is rather that α's inhabitants *can be bijected* onto β's such that every α-inhabitant is better off in α than her image (i.e. the β-inhabitant assigned to her) in β – see the bijection indicated by the arrows.

Put it this way: to say that it makes no difference who plays which part is to say that the moral judgement is immune to permutations of the individuals – immune, that is, to the swapping of roles. Thus, if Pareto-superiority suffices to establish betterness, then so does Pareto-superiority after role-swapping.[27]

---

[27] As to permutations and (their role in) standard explications of universalizability, see e.g. May (1952), Rabinowicz (1979), Dasgupta (1988), sect. 1, and their sources. Pareto-superiority and universalizability were first married in Suppes (1966), see also Kern/Nida-Rümelin

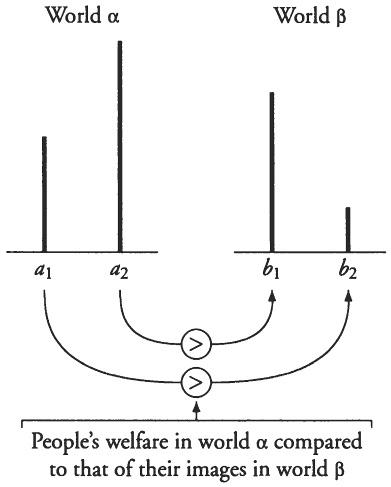
World α
World β
a_1
a_2
b_1
b_2
>
>
People's welfare in world α compared to that of their images in world β

Fig. 1: Pareto and universalizability

These thoughts would seem to suggest a principle of the following type:

> Let α and β be any two possible worlds; if there is a bijection of the set of possible individuals onto itself such that every possible individual is at least as well off with world α as their image is with world β, and at least one of them is even better off, then world α is better than world β.

Suppose, however, we came across a possible world α (with inhabitants $\dots a_{-2}, a_{-1}, a_0, a_1, a_2 \dots$) and a possible world β (with inhabitants $\dots b_{-2}, b_{-1}, b_0, b_1, b_2 \dots$) in which welfare is distributed over the denumerably infinite number of inhabitants as follows:

$$\text{World } \alpha \left[\begin{array}{llllllllll} \text{individual } a_i\text{:} \dots a_{-4} & a_{-3} & a_{-2} & a_{-1} & a_0 & a_1 & a_2 & a_3 & a_4 \dots \\ u_{a_i}(\alpha)\text{:} \quad \dots -\frac{4}{5} & -\frac{3}{4} & -\frac{2}{3} & -\frac{1}{2} & 0 & \frac{1}{2} & \frac{2}{3} & \frac{3}{4} & \frac{4}{5} \dots \end{array} \right.$$

$$\text{World } \beta \left[\begin{array}{llllllllll} u_{b_i}(\beta)\text{:} \quad \dots -\frac{5}{6} & -\frac{4}{5} & -\frac{3}{4} & -\frac{2}{3} & -\frac{1}{2} & 0 & \frac{1}{2} & \frac{2}{3} & \frac{3}{4} \dots \\ \text{individual } b_i\text{:} \dots b_{-4} & b_{-3} & b_{-2} & b_{-1} & b_0 & b_1 & b_2 & b_3 & b_4 \dots \end{array} \right.$$

The current version of our principle would entail both that α is better than β (see for instance the bijection $f(a_i) = b_i$) and that β is better than α (see for instance the bijection $f(a_i) = b_{i+2}$). The apparent moral force of one bijection

---

(1994), sects. 8.3.2 and 9.2.

is opposed by that of the other, and jointly they generate a deontic contradiction.[28] Seeing the problem is seeing the solution it calls for: the positive condition, viz. the existence of a certain bijection, should be said to suffice only if we cannot find a 'morally reverse' bijection as well. Adding this negative condition, we get

The Format of a General Universal Pareto Principle (FGUPP):

> Let Ω be the set of all possible individuals; let α and β be possible worlds.
>
> $$\begin{array}{lll} \text{(i)} & \exists \text{ bijection } \mathrm{f}: \Omega \to \Omega & : \quad \forall x \in \Omega\ u_x(\alpha) \geq u_{\mathrm{f}(x)}(\beta)\,\& \\ & & \quad \exists x \in \Omega\ u_x(\alpha) > u_{\mathrm{f}(x)}(\beta), \text{ and} \\ & \neg\exists \text{ bijection } \mathrm{f}: \Omega \to \Omega & : \quad \forall x \in \Omega\ u_x(\beta) \geq u_{\mathrm{f}(x)}(\alpha) \\ & \Rightarrow \alpha >_{\heartsuit} \beta & \end{array}$$
>
> $$\begin{array}{lll} \text{(ii)} & \exists \text{ bijection } \mathrm{f}: \Omega \to \Omega & : \quad \forall x \in \Omega\ u_x(\alpha) \geq u_{\mathrm{f}(x)}(\beta), \text{ and} \\ & \neg\exists \text{ bijection } \mathrm{f}: \Omega \to \Omega & : \quad \forall x \in \Omega\ u_x(\beta) \geq u_{\mathrm{f}(x)}(\alpha)\,\& \\ & & \quad \exists x \in \Omega\ u_x(\beta) > u_{\mathrm{f}(x)}(\alpha) \\ & \Rightarrow \alpha \geq_{\heartsuit} \beta & \end{array}$$
>
> [In prose:
>
> (i) If there is a bijection of the set of possible individuals onto itself such that every possible individual is at least as well off with world α as their image is with world β, and one of them even better off, and there is no bijection that leaves every possible individual at least as well off with β as their image with α, then α is better than β.
>
> (ii) And, as usual, the slightly weakened consequent for the slightly weakened antecedent.]

So much for the Format of a General Universal Pareto Principle, FGUPP. Its only moral ingredients are Pareto-superiority and universalizability. FGUPP is the format of the strongest claim that possibilists can make without beginning to genuinely weigh welfare levels against each other.

We should briefly address an ontological worry, too. Talk of "possible preferrers", "all possible people", and the like might be objected to on metaphysical grounds. Perhaps, as Quine is well known to believe, the

---

[28] I'm grateful to Uwe Bombosch for pointing this out to me. The principle would be fine if at least one of the worlds had the following properties: the number of different welfare levels in it is finite, and no more than one welfare level occurs an infinite number of times. Both parts of this condition are fulfilled, for example, if at least one of the worlds has a finite population.

> "slum of possibles is a breeding ground for disorderly elements. Take, for instance, the possible fat man in that doorway; and again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones?"[29]

However, the jargon that is so offensive to some ears can be eliminated from our principles. There is an equivalent way of stating FGUPP (and hence GUPP itself, to be met in section 6) that speaks, for every two possible worlds α and β, just of the people that exist in α and of those that exist in β, but at no stage of "all possible people". Even more radically, we could stick to *predicates* of the type "is a preferrer and leads a life of utility so-and-so", count, in the possible worlds at issue, the instances (no matter 'who' they are) of each of these predicates, and restate FGUPP in those terms, i.e. in terms of the cardinality of the extensions of the predicates. By then, we will have avoided both the reference to "all possible people" and the minefield of trans-world identity. If, however, your objections go further still, and concern possible *worlds*, then you are likely to have a hard time doing ethics at all.

6. *The General Universal Pareto Principle (GUPP), the Mere Addition Paradox, and the Repugnant (and the Very Repugnant) Conclusion*

As the "F" in its name suggests, our latest result, FGUPP, is still only the format of a principle. Every possibilist should accept it. To get from the format to the real thing, we need a component telling us *which* utility level is to be assigned to non-existence. Here, possibilists will part company, as was pointed out at the beginning of the previous section. In sections 1 to 4, however, we had argued that the component should be antifrustrationist. If that is so, then the right way to get rid of the "F" is simply to add to FGUPP the Principle of Antifrustrationism, PAF, from section 4. We thus get

The General Universal Pareto Principle (GUPP):

> FGUPP and PAF.[30]

GUPP, then, is a Pareto principle enriched by universalizability and antifrustrationism. Having constructed and justified it step by step, we can turn to some of its properties and consequences. Section 7 will recite a whole list of

---

[29] Quine (1948), p. 4; see also Quine (1960), §§ 8 and 50.

[30] For FGUPP and PAF see above, pp. 529 and 524 respectively.

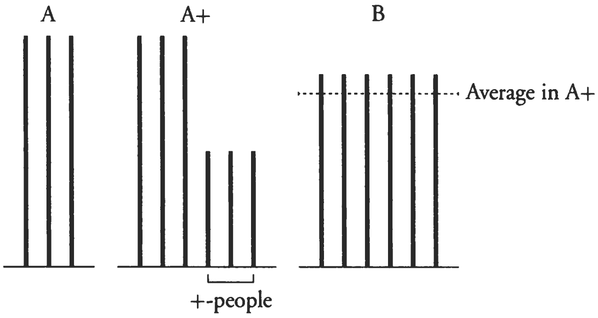

Fig. 2: The Mere Addition Paradox

them; before that, here comes GUPP's comment on two notorious cases of different people choices.

The Mere Addition Paradox

First, there is Derek Parfit's Mere Addition Paradox.[31] In the three diagrams from figure 2, let, for some large k (say ten billion or more), every vertical dash represent k people. Apart from this, the diagrams work as before (see sect. 5).

World A has $3k$ inhabitants who are very very happy, all of them equally so. (As usual, we assume "happy" and similar terms to be translatable into preference jargon.) We call their level of happiness the A-level, people on the A-level we call A-people.

World A+ has twice as many inhabitants as A; half of them are as happy as their colleagues from A, the other half a little less so, but still very happy. In a sense, A+ can be obtained by just adding happy people to A. (Hence the "+" and the name of the paradox.) The expressions "+-people" and "+-level" refer to the 'added' beings and their happiness.

World B, too, has $6k$ inhabitants, just as many as A+, that is; all of them are equally happy, less so than the A-people, but more so than the +-people, and, luckily for them, closer to the A-level than to the +-level. The average welfare from A+ is drawn into B's diagram as a dotted line.

Suppose you reason as many people do when first confronted with pairwise comparisons of these worlds:

(P 1) $A >_{\heartsuit} B$

---

[31] See Parfit (1984), ch. 19.

P 1 says that world A is better than world B.[32] This is a straightforward application of Narveson's slogan that we have obligations to make people happy, but no obligations to make happy people.[33] People in B are less happy than those in A, and that makes, ceteris paribus, B worse than A. The only conceivable candidate for a feature that might outweigh B's worseness is that in B the number of people, and therefore the sum total of happiness, is higher. But Narveson's slogan tells us that the candidate fails: happy or not, *raising* the number of people is of no value. Hence there is nothing to counterbalance B's ceteris paribus worseness, hence B is worse than A.

$$(\text{P 2}) \qquad A+ \geq_{\heartsuit} A$$

P 2 says that world A+ is at least as good as world A. The claim is based on the idea that, other things being equal, it can hardly be wrong to bring very happy people (like the +-lot) into existence. They will be grateful for being alive and will enjoy it. And in the case at hand, other things *are* equal: adding the +-people makes nobody else worse off. Note that we are only saying "at least as good as", and reserve judgement with respect to "better than". We are presenting things so as not to leave out the Narvesonians. They would deny the "better than" claim, since they take adding the +-people to be of no value. So all that P 2 says is that adding them does no harm. It doesn't say that adding them is good.

$$(\text{P 3}) \qquad B >_{\heartsuit} A+$$

P 3 says that world B is better than world A+. This verdict seems morally overdetermined: B has a higher average and total utility than A+, and the distribution is perfectly equal. Now, if of two worlds with the same number of inhabitants one world contains *more* welfare than the other *and* distributes it equally, whereas the other one doesn't, then how can it fail to be better?

By this time, however, you have contradicted yourself – at least if we assume, as usual, "better than" to be transitive:

$$\left.\begin{array}{ll} (\text{P 3}) & B >_{\heartsuit} A+ \\ (\text{P 2}) & A+ \geq_{\heartsuit} A \end{array}\right] \Rightarrow \; B >_{\heartsuit} A \quad \lightning \quad (\text{P 1}) \quad A >_{\heartsuit} B$$

It follows from P 3 and P 2 that B is better than A, which is denied by P 1. Confronted with this contradiction, many people have a hard time deciding which of the Ps to withdraw. The indecision is mirrored in the literature. Each

---

[32] A view that was first expressed by C. D. Broad and David Ross, see note 35 below.

[33] Cf. note 21.

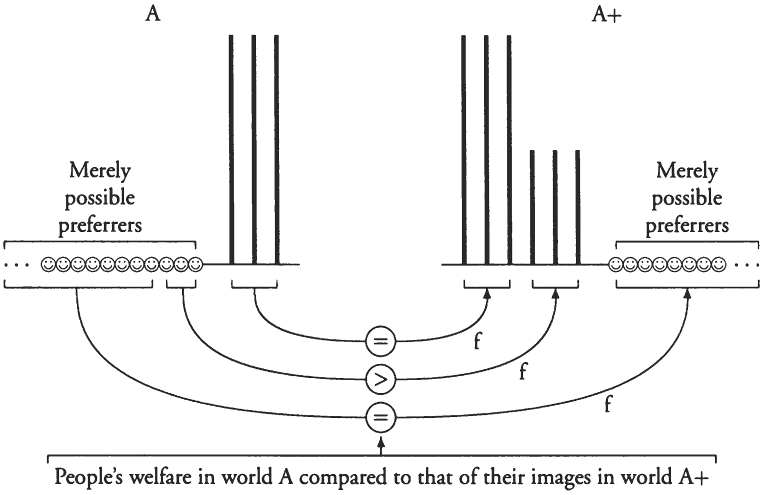
A
A+
Merely possible preferrers
Merely possible preferrers
f
f
f
=
>
=
People's welfare in world A compared to that of their images in world A+

Fig. 3: GUPP and the Mere Addition Paradox

of the three premisses, as well as the transitivity of betterness, has been rebuffed in attempts to dissolve the paradox.[34]

GUPP meets the Mere Addition Paradox

What does GUPP have to say about the Mere Addition Paradox? Ex hypothesi, the +-people each enjoy less than maximum welfare. It follows (see sect. 4 above) that each of them has at least one frustrated preference. This in turn implies (see sect. 3 above) that it is wrong to bring them into existence. Therefore, premiss P 2 is wrong, and world A+ is worse than world A. If we drag in GUPP's full machinery, and paint in the mere possibilia (with one smiley representing *k* of them), a bijection like f, shown in figure 3, will settle the issue.

Bijection f maps A's preferrers onto the happier half of those in A+; of A's non-preferrers, f maps 3*k* onto the +-people from A+, and the rest onto the non-preferrers from A+. Clearly, there is no 'morally reverse' bijection, since we shall not find anybody of whom it is true that the +-people are at least as well off with world A+ as that person is with world A.

---

[34] Many of these attempts are documented in part IV of Parfit (1984). A distant relative to GUPP's solution could be implicit in Tooley's views, cf. principle S in sect. 7.43 of his (1983). But it is hard to tell, and his principles T_1 and T_2 from the same section suggest a different outlook; see also Parfit (1984), sect. 124.

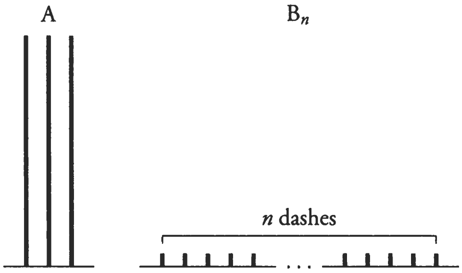

Fig. 4: The Repugnant Conclusion

The Repugnant and the Very Repugnant Conclusion

Next, and equally notorious, what the literature knows as the Repugnant Conclusion.[35] In the diagrams from figure 4, let again every vertical dash represent some large number k of people ($k \geq$ ten billion).

World A has $3k$ inhabitants who are all very very happy. A world of the type B_n has n times k inhabitants who all consider their lives barely worth living; they are so close to the brink that, if their life contained just one more stomach-ache than it does, they would prefer not to have been born.

Now, total utilitarianism is the view that we always ought to maximize the total amount of utility. It is generally thought that total utilitarianism implies

The Repugnant Conclusion:

> There is an n such that a world of type B_n is better than world A.

For any number of very very happy preferrers, says the Repugnant Conclusion, we can imagine a larger population whose existence would be better, though its members have lives they consider just barely worth living. This is counterintuitive.

There seems to be more trouble ahead for total utilitarians. Once they assign some positive value, however small, to the creation of each person who has a weak preference for leading *her* life rather than no life, then how can they stop short of saying that some large number of such lives can compensate

---

[35] See Parfit (1984), ch. 17. The Repugnant Conclusion is closely related to the Mere Addition Paradox; for this connection, see ibid., ch. 19. The underlying problem (see also Birnbacher 1988, sects. 2.3.2 f.) is that average and total utility can point in different directions – a discovery that marked the beginning of population ethics as we now know it: Sidgwick (1874), book IV, ch. I, § 2. The utilitarian recommendation to increase total even at the cost of average utility was first adduced as an objection to utilitarianism by Broad (1930, pp. 249 f.) and Ross (1939, pp. 69 f.).

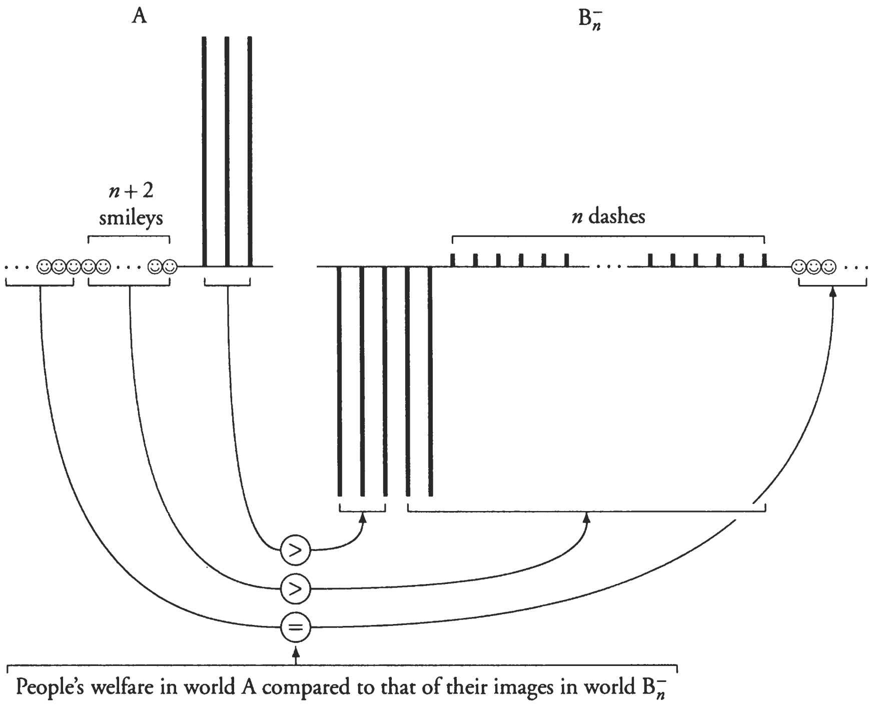
A
B_n^-
$n+2$ smileys
n dashes
>
>
=
People's welfare in world A compared to that of their images in world B_n^-

Fig. 5: GUPP and the Very Repugnant Conclusion

for the creation of lots of *dreadful* lives, lives in pain and torture that nobody would want to live? The problem is illustrated in figure 5 (where again every smiley stands for k possibilia): if we add to a world B_n billions of terrible lives, we get a world B_n^- (read: "B_n minus") that we can compare to world A. Total utilitarianism, it would seem, is committed to what we could call

The Very Repugnant Conclusion:

> There is an n such that a world of type B_n^- is better than world A.

GUPP, the General Universal Pareto Principle, denies both the Repugnant and the Very Repugnant Conclusion; for the latter, a sketch of a relevant bijection is included in figure 5. Any bijection that maps A's inhabitants on inhabitants of B_n^- will do.

Now, if total utilitarianism entails, whereas GUPP rejects, the Repugnant Conclusion, are the two doctrines incompatible? Yes, if. But, contrary to received opinion, total utilitarianism does *not* entail the Repugnant Conclusion – not if utility is, as section 1 argued, a measure of avoided preference frustration. Utilitarianism, thus read, asks us to minimize the amount of preference frustration, and there is less frustration in world A than in worlds of the type B_n^- or, for $n \geq 3$, of the type B_n. Antifrustrationism liberates total utilitarianism from the Repugnant and the Very Repugnant Conclusion. (Side-effect: John Rawls loses his only argument to the effect that total utilitarianism would not be chosen behind his 'veil of ignorance'; the argument is in sect. 27 of his 1971.)

However, other than most theories of different people choices, antifrustrationism is not *based* on, nor motivated by, the premiss that the Repugnant Conclusion is repugnant. It is based on general reflections about the concept of welfare (see sect. 1), and it denies the Repugnant Conclusion without saying or assuming either that the Repugnant Conclusion is, or that it is important to avoid aggregational judgements that are, counter-intuitive.

7. *GUPP: Selected Properties and Consequences*

In essence, different people choices should be looked at as what we can call *different preferences choices* – as choices that make a difference to the number or identity of *preferences* that will ever exist. This paper started mongering a theory of the latter, and applying it to the former.

The route we have travelled so far is this. Section 1 showed (at some length) that what counts is not the existence of satisfied preferences, but the non-existence of frustrated ones, a doctrine we called antifrustrationism. Section 2 added to antifrustrationism the spirit of Pareto; the result was a Pareto principle on the somewhat unusual level not of people but of preferences: the Principle of Pareto-Superiority among Wishes (POPSAW). Section 3 presented and discussed some moral judgements that POPSAW entails about the existence of preferrers. Section 4 translated antifrustrationism from preference jargon into utility jargon, a translation achieved by the Principle of Antifrustrationism (PAF); PAF was shown to be a possibilistic doctrine: it can be looked at, that is, as assigning a welfare level even to non-existence. Section 5 addressed itself to every possibilist, antifrustrationist or not; the question was how much *all* possibilist moralities, regardless of their differences with respect to the value of non-existence, can say before touching on truly compensatory issues. The answer was the Format of a General Universal Pareto Principle (FGUPP). FGUPP is Suppes's good old conjunction of Pareto-superiority and universalizability, adapted to meet the extra requirements of different people choices and of infinite populations. Section 6 linked FGUPP to antifrustrationism (more precisely, to PAF, from sect. 4), yielding the General Universal Pareto Principle (GUPP). GUPP is, as far as moral principles are concerned, all I want to sell today. It was pointed out what the principle has to say about the Mere Addition Paradox and the Repugnant Conclusion.

By way of conclusion, the present section is about to end the paper with an unordered list of some of GUPP's properties and implications.

GUPP is blatantly similar to, and entails, the Ordinary Pareto Principle OPP. However, GUPP is, even if restricted to same people choices, not equivalent to, but stronger than OPP.[36] If you care for an equation:

$$\text{GUPP} = \text{Pareto} + \text{universalizability} + \text{antifrustrationism}.$$

GUPP itself is incomplete (there are worlds it does not compare to each other), because it wants to stay aloof of most controversies in preference-based ethics. What it does treat are the basic issues in the morality of possible preferences and possible people. It does so mainly by inquiring into the *nature* of welfare. When the time comes to ask what we should do *about* welfare, all that our principle endorses are universalizability and ordinary Pareto-superiority. Thus, *GUPP is compatible with, and can be plugged into, practically every social welfare function, utilitarian or not, that anybody would ever dream of defending.*

What does GUPP make of procreation? A complete and precise answer could hardly be simpler than GUPP itself, but we can point out a few interesting consequences. GUPP entails POPSAW (the Principle of Pareto-Superiority among Wishes, presented in sect. 2), and hence the implications of POPSAW that were discussed in section 3: that empty worlds are at least as good as any other worlds, and better than our world; and that it is ceteris paribus wrong to bring people into existence who will have an unfulfilled preference.

But "ceteris paribus" aside – what about real-life cases? What about parents whose happiness is increased by the creation of a person who will have unfulfilled wishes? How does GUPP 'weigh' the intrinsic badness of the creation against the parents' benefit?[37] Figure 6 might be instructive. Let each of the dashes represent one actual person, each of the smileys one possible preferrer. The story behind the picture is this. Individual a is alone and wants a

---

[36] Figure 1, from p. 528, can illustrate the point. Pretend that b_1 is the same person as a_1, and b_2 the same as a_2; the choice between α and β is then a same people choice. a_1 is better off in β than in α, a_2 vice versa. OPP refuses to 'weigh' these pros and cons against each other and thus cannot rank the worlds. GUPP, however, universalizes the problem away and finds α better than β.

[37] "weigh" in scare quotes, since GUPP doesn't engage in any real weighing – in no weighing, that is, after identity has been bijected away; see sect. 5 above.

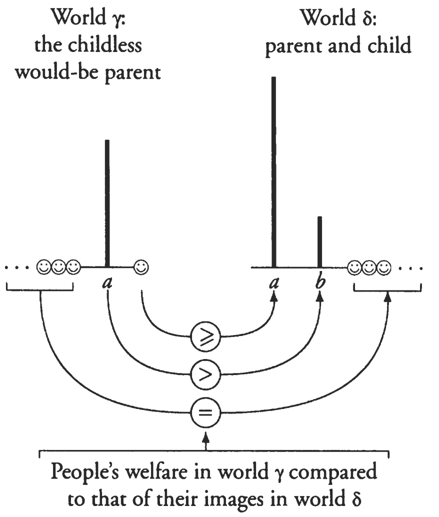
World γ:
the childless
would-be parent
World δ:
parent and child
a
a
b
≥
>
=
People's welfare in world γ compared
to that of their images in world δ

Fig. 6: GUPP and a possible child

child. In world γ, his wish is frustrated and he remains alone; in world δ, his wish is fulfilled and he has the desired heir, *b*. As is indicated in figure 6, any bijection that maps *a* onto *b* shows us that GUPP declares γ to be better than δ. The child ought not to be conceived.

But if we modify the case and make *b*'s dash in world δ higher than *a*'s in world γ, then no bijection of the type mentioned in GUPP's antecedents can be found, and GUPP is silent on the comparative value of γ and δ. Everything will then depend on how we complete GUPP, in other words: on our *general* creed about the aggregation of welfare. (Is quantity all that matters? What about equality? And minimum levels? Etc.) Since millions of real-life situations look like the modified case (the child would be happier than the childless would-be parent would have been), GUPP is miles away from anything like a general prohibition on real-life procreation.

GUPP entails that it is obligatory not to bring into existence an unhappy person, but not obligatory to bring into existence a happy person. GUPP also entails the Narveson type of slogan that we have obligations to make people happy (preferrers satisfied), but no obligations to make happy people (satisfied preferrers). These "asymmetries", as they have been called, have their home on the level of preferences, or Good Sentences: the existence of a frustrated preference is bad, but the existence of a satisfied preference is not good.[38]

---

[38] See sect. 1. For this explanation of asymmetry, see also Peter Singer's remarks referred to in

GUPP implies that abortion poses no intrinsic moral problems as long as the fetus has no preferences. Genetic engineering and experiments on embryos are, as far as the individual in question is concerned, not worse than alternative actions as long as they do not inflict more preference frustration on that individual.

GUPP's antifrustrationism also affects the question whether preference-based ethics should concentrate on people's *global preferences*.[39] Derek Parfit thinks so, but his crucial thought experiment assumes that giving equal weight to non-global preferences would involve assigning positive value to satisfied extra preferences. That is precisely what GUPP's component PAF bids us not to do, and it is unclear whether antifrustrationist mutations of Parfit's scenario would make the equal treatment of global and local preferences seem worth rejecting.

GUPP entails the uncontroversial readings of what Parfit calls the

> "Same Number Quality Claim, or Q:
>
> If in either of two possible outcomes the same number of people would ever live, it would be worse if those who live are worse off, or have a lower quality of life, than those who would have lived."[40]

If this referred to the sum of the people's quality of life, then Q would be strong and debatable. It would endorse what one might call same number utilitarianism – utilitarianism, that is, for all "same number choices"[41]. GUPP is much weaker. It is compatible with, but not committed to, same number utilitarianism. (It is not committed to any form of utilitarianism.) There are, however, less debatable readings of Q (Pareto-readings, so to say). The clause that "those who live are worse off [...] than those who would have lived" could mean that everybody who lives is worse off than everybody who would have lived; or, more generally, that those who live can be bijected onto those who would have lived such that every actual individual is worse off than the possible one assigned to it. For both these readings, GUPP entails Q; and for both of them, the task, set to us by Parfit, to say "how Q should be justified, or more fully explained"[42], is manageable (see sect. 5 above): Q, thus read, follows from the conjunction of Pareto's principle and universalizability.

---

note 2 above; as to Narveson's slogan, cf. note 21 above.

[39] See above, p. 514 and note 13.

[40] Parfit (1984), p. 360.

[41] Parfit's term (1984, sect. 120) for choices between outcomes that contain the same number of preferrers.

[42] Parfit (1984), p. 361.

GUPP – since it entails Q – also entails that a woman ought to bring into existence, if anybody, then a happier preferrer rather than an unhappier one. If this involves avoiding the existence of the happy preferrer, then, ceteris paribus, she ought to do it. If this avoidance involves killing (and replacing) a preference-less fetus, then, ceteris paribus, she still ought to do it. This has been called the *Non-Identity Problem*[43], for it involves something like avoiding one person's existence 'for the sake of' somebody else's.

GUPP contradicts the so-called *Absurd Conclusion* that, given a pattern of welfare dashes, it can make a moral difference *when* the respective lives are lived.[44] GUPP just has no time parameter: it is blind to the question *when* a welfare dash of a certain height is realized. According to GUPP, two worlds that realize the same pattern of dashes – no matter when – are of equal value.

GUPP, as has been shown in the previous section, solves the Mere Addition Paradox and avoids, and even teaches total utilitarianism how to avoid, the Repugnant and the Very Repugnant Conclusions.

We can thus nominate GUPP for the position of *Theory X*, or, at any rate, for that of its hard core. To explain: actions that make a difference to the number or identity of people who will come into existence raise moral puzzles, some of which have been mentioned above. Derek Parfit has been most active in putting them on the agenda and most ingenious in discussing them. In his *Reasons and Persons* he considers a range of these issues. Each of them is vexing in itself, or it annoys a tempting or common theory of beneficence, or it adds a requirement (innocent enough on its own, perhaps) to a *list* of criteria all of which an acceptable theory of beneficence should meet. Parfit's work has triggered off a massive hunt for such a theory – for "Theory X", as he himself baptized the desideratum. "As I argued," Parfit concludes,

> "we need a new theory of beneficence. This must solve the Non-Identity Problem, avoid the Repugnant and Absurd Conclusions and solve the Mere Addition Paradox. I failed to find a theory that can meet these four requirements."[45]

GUPP meets them.

---

[43] Parfit (1984), p. 359; see also Kavka (1982).

[44] See Parfit (1984), pp. 410 f.

[45] Parfit (1984), p. 443.

Acknowledgements and References

I'm grateful to Ulla Wessels for countless exchanges; many of the arguments in her and my work on possible preferences are ours rather than either hers or mine. I profited from discussing various draft versions, or the ideas in them, with, among others, Christopher Abbey, Krister Byqvist, Sven Danielsson, Wulf Gaertner, Richard Hare, Wolfgang Lenzen, Seán Matthews, Georg Meggle, Elijah Millgram, Christian Munthe, Yew-Kwang Ng, Björn Petersson, Peter Ryman, Peter Singer, Rysiek Sliwinsky, and Harald Stücker, all of whom came up with important observations. Many thanks to more than a dozen audiences for listening to, and discussing, prototypes of this paper. And to everybody involved in running the University of Pittsburgh's Center for Philosophy of Science, where, in 1992, an early version of this paper was conceived. The bulk of the work was part of the research project "Was zählt?"; I am grateful to the Deutsche Forschungsgemeinschaft (DFG) for its support of this project.

An outline of the ancestors and near and distant relatives of antifrustrationism will have to await another occasion. (See, however, the sources listed in notes 2 and 21.) It is instructive, for example, to compare the doctrine to the relevant teachings of Buddha, the Stoics, Schopenhauer, or Albert Ellis, to Seana Shiffrin's recent work, to pessimism (in the various meanings of that word), and to what has become known as 'negative utilitarianism'. Some similarities notwithstanding, all of these differ from antifrustrationism in important respects.

Kappel (1996) is, and Danielsson (1996) and Wolf (1997) contain, an instructive critique of this paper. I hope to respond to them on some other occasion. As to Kappel's paper, three brief caveats: (i) Kappel's introduction lumps together antifrustrationists and prophylacticists, which is misleading; I do not believe that preferring is painful or bad. (ii) The desire to avoid certain notorious judgements in population ethics is not a major reason why I embrace antifrustrationism. (iii) Other than in the early draft that Kappel criticizes, Narveson's slogan is no longer a premiss of my argument.

BENTHAM (1825). Jeremy Bentham: *The Rationale of Reward*, in vol. II of J. Bowring (ed.): *The Works of Jeremy Bentham*, Edinburgh 1838–43; *Rationale* first publ. in 1825.

BIRNBACHER (1988). Dieter Birnbacher: *Verantwortung für zukünftige Generationen*, Stuttgart 1988.

BRICKER (1980). Phillip Bricker: "Prudence", *Journal of Philosophy* 77 (1980).

BROAD (1930). C. D. Broad: *Five Types of Ethical Theory*, London 1930.

DANIELSSON (1996). Sven Danielsson: "Between the Repugnant and the Gloomy? Some Preferentialisms", in Lindström et al. (1996).

DASGUPTA (1988). Partha Dasgupta: "Lives and Well-Being", *Social Choice and Welfare* 5 (1988).

FEHIGE/WESSELS (1998). Christoph Fehige and Ulla Wessels: "Introduction to Possible Preferences", this volume.

FOTION/HELLER (1997). N. Fotion and Jan C. Heller (eds.): *Contingent Future Persons*, Dordrecht 1997.

HARE (1975). Richard M. Hare: "Abortion and the Golden Rule", in Hare (1993); paper first publ. in 1975.

HARE (1993). Richard M. Hare: *Essays on Bioethics*, Oxford 1993.

HARE (1998). Richard M. Hare: "Preferences of Possible People", this volume.

KANT (1785). Immanuel Kant: *Grundlegung zur Metyphysik der Sitten*, Berlin 1911, vol. IV of the Royal Prussian Academy edition of *Kant's gesammelte Schriften*; *Grundlegung* first publ. in 1785.

KAPPEL (1996). Klemens Kappel: "On Frustrationism and Satisfactionism", in Rabinowicz (1996).

KAVKA (1982). Gregory Kavka: "The Paradox of Future Individuals", *Philosophy and Public Affairs* 11 (1982).

KERN/NIDA-RÜMELIN (1994). Lucian Kern and Julian Nida-Rümelin: *Logik kollektiver Entscheidungen*, Munich 1994.

LENZEN (1998). Wolfgang Lenzen: "Who Counts?", this volume.

LINDSTRÖM ET AL. (1996). Sten Lindström, Rysiek Sliwinski, and Jan Österberg (eds.): *Odds and Ends*, Uppsala 1996.

LOCKWOOD (1979). Michael Lockwood: "Singer on Killing and the Preference for Life", *Inquiry* 22 (1979).

MAY (1952). Kenneth O. May: "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision", *Econometrica* 20 (1952).

MCMAHAN (1998). Jeff McMahan: "Preferences, Death, and the Ethics of Killing", this volume.

NARVESON (1967A). Jan Narveson: *Morality and Utility*, Baltimore 1967.

NARVESON (1967B). Jan Narveson: "Utilitarianism and New Generations", *Mind* 76 (1967).

PARETO (1906). Vilfredo Pareto: *Manuale di Economia Politica*, Milan 1906.

PARFIT (1984). Derek Parfit: *Reasons and Persons*, Oxford 1987; first ed. first publ. in 1984.

PETERSSON (1993). Björn Petersson: "Den moraliska asymmetrin mellan lycka och lidande", *Filosofisk Tidskrift* 14 (1993).

QUINE (1948). Willard V. O. Quine: "On What There Is", in Quine (1953); paper first publ. in 1948.

QUINE (1953). Willard V. O. Quine: *From a Logical Point of View*, second ed., Cambridge, Mass., 1961; first ed. first publ. in 1953.

QUINE (1960). Willard V. O. Quine: *Word and Object*, Cambridge, Mass., 1960.

RABINOWICZ (1979). Włodzimierz Rabinowicz: *Universalizability*, Dordrecht 1979.

RABINOWICZ (1996). Wlodek Rabinowicz (ed.): *Preference and Value*, Lund 1996.

RAWLS (1971). John Rawls: *A Theory of Justice*, Cambridge, Mass., 1971.

ROSS (1939). W. David Ross: *Foundations of Ethics*, Oxford 1939.

SEN (1970). Amartya K. Sen: *Collective Choice and Social Welfare*, San Francisco 1970.

SIDGWICK (1874). Henry Sidgwick: *The Methods of Ethics*, first ed., London 1874.

SIDGWICK (1907). Henry Sidgwick: *The Methods of Ethics*, seventh ed., London 1907; first ed. first publ. in 1874.

SINGER (1980). Peter Singer: "Right to Life?", letter in the *New York Review of Books*, August 14, 1980.

SINGER (1993). Peter Singer: *Practical Ethics*, second ed., Cambridge 1993; first ed. first publ. in 1979.

SINGER (1998). Peter Singer: "Possible Preferences", this volume.

SPRIGGE (1968). Timothy L. Sprigge: "Professor Narveson's Utilitarianism", *Inquiry* 11 (1968).

SUPPES (1966). Patrick Suppes: "Some Formal Models of Grading Principles", *Synthese* 16 (1966).

TOOLEY (1983). Michael Tooley: *Abortion and Infanticide*, Oxford 1983.

VETTER (1969). Hermann Vetter: "The Production of Children as a Problem for Utilitarian Ethics", *Inquiry* 12 (1969).

VETTER (1971). Hermann Vetter: "Utilitarianism and New Generations", *Mind* 80 (1971).

WESSELS (1998). Ulla Wessels: "Procreation", this volume.

WOLF (1997). Clark Wolf: "Person-Affecting Utilitarianism and Population Policy", in Fotion/Heller (1997).