

Derek Parfit's Contributions to Philosophy

by

WLODEK RABINOWICZ

Lund University and London School of Economics

IN 2014, THE ROYAL SWEDISH Academy of Sciences awarded Derek Parfit the Rolf Schock Prize in Logic and Philosophy. In its motivation, the Academy stressed Parfit's ground-breaking contributions to theory of personal identity, population ethics and analysis of the structure of moral theories.

The list of philosophers and logicians who have received the Rolf Schock Prize is as yet relatively short. It starts with Willard v. Quine (1993) and continues with Michael Dummett (1995), Dana Scott (1997), John Rawls (1999), Saul Kripke (2001), Solomon Feferman (2003); Jaakko Hintikka (2005), Thomas Nagel (2008), Hilary Putnam (2011) and Parfit (2014). The prize was originally awarded every second year, but since 2005 it has been given every third year. There are also Rolf Schock prizes in Mathematics, Music and Visual Arts.

This special issue of *Theoria* is devoted to Derek Parfit's contributions to philosophy. It contains the talk on population ethics he gave at the prize symposium in Stockholm in October 2014. In addition, it includes papers written by the four other contributors to the symposium: Gustaf Arrhenius, Ruth Chang, Ingmar Persson and Larry Temkin.

Derek Parfit was born in 1942 in China, where his parents – both of whom were physicians – gave courses in preventive medicine at mission-run hospitals. He went to Eton and then, in 1961, started reading history at Balliol College in Oxford. In 1965–1966 he went on to Columbia and Harvard as a Harkness Fellow. It was then that he became fascinated by philosophy. Upon his return to Oxford he won a seven-year Prize Fellowship at All Souls College. He remained at the same college even thereafter, as a research fellow; indeed, he spent his entire academic career at All Souls, where he presently (since 2010) is Emeritus Senior Research Fellow. He has been a visiting professor at several North American universities: Harvard, NYU and Rutgers. Some of these trans-Atlantic assignments he retained even after retirement from Oxford. He is a member of the British Academy and of the American Academy of Arts and Sciences.

Parfit's most famous and influential work, *Reasons and Persons*, was published in 1984. In its importance for the development of moral philosophy it is often compared to Henry Sidgwick's (1981 [1907]) great nineteenth-century classic *The Methods of Ethics*. It opened new and exciting vistas and radically changed the landscape in which moral philosophy was pursued. In 2011, Parfit published his

second magnum opus, *On What Matters*, in two volumes comprising altogether nearly 1,400 pages. Even this book has been compared to Sidgwick's *Methods*. Its second volume includes other philosophers' objections to the theory presented in the first volume, together with Parfit's replies. A third volume is under way, with further replies, this time to the objections included in a forthcoming collection *Does Anything Really Matter?*, edited by Peter Singer.

Apart from these two monumental works, Parfit published a large number of important papers, among which one should mention "Personal Identity" (1971), "Is Common-Sense Morality Self-Defeating?" (1979), "Future Generations: Further Problems" (1982), "Rationality and Time" (1983), "The Unimportance of Identity" (1995), "Equality and Priority" (1997a), "Reasons and Motivation" (1997b), "Rationality and Reasons" (2001), "Justifiability to Each Person" (2003), "Overpopulation and The Quality of Life" (2004a), "What We Could Rationally Will" (2004b), "Normativity" (2006) and "Another Defence of the Priority View" (2012). Some of the papers have been incorporated into his books; some further develop topics from *Reasons and Persons*; while others take up new subjects.

With the publication of *Reasons and Persons* in the mid-1980s, Derek Parfit set the direction for the mainstream of moral theory. The book's impact was momentous, not only because of the ideas it put forward, but also due to its philosophical style, with a frequent usage of hypothetical examples appealing to the reader's moral intuitions. *Reasons and Persons* consists of four parts. In the first part, "Self-Defeating Theories", the focus is on structural analysis of normative theories. Three such theories are subjected to closer investigation: (i) self-interest theory, which prescribes that one maximizes the satisfaction of one's own interests over time, (ii) consequentialism, which prescribes the realization of the best outcome, whether or not that outcome is best for oneself, and (iii) common-sense morality, according to which we have special obligations towards particular persons: towards our near and dear.

Parfit considers under what conditions compliance with a normative theory or, alternatively, conscious efforts to comply may defeat the very objectives the theory explicitly or implicitly puts forward. In the former case, when defeat is caused by compliance, the normative theory is directly self-defeating, while in the latter case, when defeat is caused by the efforts to comply, the defeat is indirect. Parfit argues that consequentialism is indirectly self-defeating. Impartial consequentialists are bad at maintaining close relationships with their family and friends. A sustained pursuit of the best outcome might therefore well lead to a loss of important values that essentially depend on personal relations. This is troublesome but need not show that consequentialism is an incorrect theory. Instead, the conclusion might be that we should abstain, on purely consequentialist grounds, from wholeheartedly embracing consequentialism.

On the other hand, self-interest theory and common-sense morality are directly self-defeating, not individually but collectively. If we all obey the prescriptions of such a theory, we fail to realize the objectives it gives to each of us. Common-sense morality requires a particular consideration for our nearest and dearest; however, if practised by everyone, it may well lead to an outcome that will be suboptimal for everyone concerned: for each of us, the interests of our near and dear will be less well satisfied than it could be otherwise, in the absence of such special consideration. The same goes for self-interest theory which prescribes exclusive attention to our self-interest: as is well-known, universal pursuit of self-interest gives rise to prisoners' dilemmas.

That common-sense morality is collectively directly self-defeating shows, according to Parfit, that this theory cannot be correct: it must be revised. The same does not, however, apply to self-interest theory. The latter is meant to codify individual rationality. As it does not purport to be a moral code – a code for the collective – it cannot be disqualified just because it is collectively self-defeating. If it is to be criticized, this has to be done in another way.

In the second part of the book, “Rationality and Time”, Parfit argues that self-interest theory is inconsistent in its treatment of persons and times: it privileges the self but it has no bias towards the present and the near future. It favours the self's preferences as compared with preferences of others, but it gives the same weight to preferences one holds at different times – in the present, in the near future, and in the more distant future. This inconsistency is unacceptable, according to Parfit: either we should be partial in both the temporal and the personal dimension, which would give us what he calls the Present-Aim Theory, or we should be partial in none, which is the solution he himself adopts. He also points out that self-interest theory is not fully consistent in its time-impartiality: as standardly formulated, self-interest theory does not give weight to the preferences held in the past. But this only makes it even more problematic.

Parfit fully realizes, of course, that orientation towards the present and the future and relative disregard for the past are quite fundamental to our human condition. Think only of all those situations in which we have wished to have some onerous tasks or unpleasant experience already behind us, rather than still awaiting us in the future. That our concerns are future-directed in this way is of course perfectly understandable and amenable to an evolutionary explanation given that the past lies outside our causal control. An explanation can also be given to our bias towards the near future as opposed to the more distant one. Still, Parfit contrasts our ordinary time-partial attitude with the time-neutral ‘philosophical’ perspective, to which he himself is drawn. The correct way to look upon temporally located events is to view them impartially – “*sub specie aeternitatis*”.

The third part of the book, “Personal Identity”, is devoted to analysis of personal identity over time. Like most other people, I am especially concerned about

what is going to happen to *myself* in the future. But such an attitude presupposes that personal identity over time is something that matters – a presupposition that Parfit endeavours to undermine. An important premise in his argument is that there is no irreducible I – no metaphysical substance that survives over time. The second step is a suggestion that what matters to me is that there will exist a person in the future that will have an appropriate relation to my present self – a psychological relation of continuity and similarity. This relation, however, cannot be equated with personal identity. In some science-fiction scenarios there might simultaneously exist several future persons that are related to me in this way: think for example of a division of the brain into two halves, each of which is later transplanted into a body of its own. None of these future individuals will then be able to claim identity with me now. If any of them is identical with me, then so are all of them, which is impossible given that they are distinct from each other. However, I would still have a special concern for their well-being. It follows, therefore, that it is not personal identity that matters, after all. If we now put science-fiction scenarios aside, it is important to point out that what matters – psychological relations of continuity and similarity – might well connect me to a multiplicity of simultaneously existing future persons, although this relationship will normally vary in degree. I therefore have reason to revise my self-centred attitudes when I realize that even other persons in the future will be psychologically related to me to a larger or smaller degree. Parfit appears to think that this fact can provide a reason to accept impartiality in our concern for future persons (rather than a reason to embrace some form of graded partiality). I do not quite see, however, how this last step follows.

The fourth part of the book, “Future Generations”, has probably been the most discussed. In the centre of the discussion stands Parfit’s Non-Identity Problem. When we ask whether future people will be benefited or harmed by our actions, the issue gets considerably complicated when we realize that the alternative actions that we have to choose between, for example the alternative ways to deal with the threat of global warming, will most certainly influence which people will exist in the future. What we do will directly or indirectly affect people’s behaviour in many different ways and thus will also affect the time at which they will conceive their children. Therefore, depending on which action we choose, it will be different children that are going to be born. If the choice has far-reaching consequences, then in the long run there will not be anyone who would exist if we made another choice instead. For no one it will be the case that his or her life would have been better (or worse) if we acted differently. So how can we say that these future people will be harmed or benefited by our present actions? A conclusion that lies close at hand is that the Non-Identity Problem again suggests that the normative importance of personal identity should be denied, though in this case it is not identity over time but across possible worlds.

But the difficulties for population ethics, i.e., ethical theory that deals with future generations, do not stop here. Whether or not personal identity across possible scenarios is normatively relevant, Parfit shows that seemingly very plausible ethical principles jointly entail what he has called the Repugnant Conclusion: the principles in question imply that a future world populated by people with a very high level of well-being is worse or at least not better than a possible world whose inhabitants lead lives that are barely worth living, provided only that the latter are sufficiently more numerous than the former. On one version of this derivation, the principles involved are just the following three: (i) the transitivity of betterness, (ii) the mere addition principle according to which adding lives worth living while at the same time keeping the well-being of already existing persons unchanged or perhaps even raising it a bit does not make the world worse, and (iii) a weak principle of equality according to which it is an improvement if we can equalize the well-being of the existing people, assuming that we do not change their number or decrease the well-being average. We need to give up some of these principles if we want to avoid the Repugnant Conclusion, but which principles in that case? And what principles should they be replaced with? What should a complete population ethics look like if it is not going to imply any ludicrous conclusions? Is there any such ethics at all? These are some of the questions that Parfit has left unanswered. It is not surprising that his book gave rise to such an intense discussion.

The second book, *On What Matters*, will here be described in a very cursory way. It has two main themes. The first is an attempt to achieve consensus among moral philosophers coming from different ethical points of departure. Parfit endeavours to show that Kantian ethics, contractualism and rule consequentialism all converge towards the same normative standpoint, provided that each of these views is interpreted and re-interpreted in a satisfactory way. Interpretation invites modification, and modifications, if reasonable, lead to convergence. To use his metaphor, these normative theories climb the same mountain, from different sides, and thus can meet at the summit. They can all agree on the Triple Theory, as Parfit calls it, according to which, roughly, “an act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable” (Parfit, 2011, vol. I, p. 413).

The second theme concerns the notion of a normative reason. Parfit is one of a number of philosophers who have in recent years managed to radically re-orient discussions about normativity. The earlier focus on *ought* as the central normative concept has nowadays given way to a focus on the concept of *a reason*. What ought to be done is determined by the balance of reasons. The second main theme of the book is a sustained defence of objectivism about reasons. As part of this overall project, Parfit criticizes several competing accounts – different subjectivist, non-cognitivist or relativistic conceptions of reasons. He argues for the existence of objective truths to the effect that such-and-such facts provide or do not provide

normative reasons for such-and-such actions or attitudes. It need not be pointed out, I am sure, that his position on this fundamental issue is highly controversial.

Reasons and Persons was a book that largely determined the direction for the mainstream of moral philosophy from the mid-1980s onwards. It led to a tidal change by posing entirely new problems and proposing highly provocative ideas: the Non-Identity Problem, the need to re-think the entire project of population ethics, the normative unimportance of personal identity, etc. Parfit's more recent book, *On What Matters*, did circulate in different manuscript versions for many years prior to its publication. Nevertheless, it still is too early to judge its influence on the ensuing course of philosophical discussions about the nature of morality. It should not be surprising therefore that the motivation for the 2014 Rolf Schock Prize focuses on Parfit's seminal contributions in *Reasons and Persons*.

References

- PARFIT, D. (1971) "Personal Identity." *Philosophical Review* 80: 3–27.
- PARFIT, D. (1979) "Is Common-Sense Morality Self-Defeating?" *Journal of Philosophy* 76(10): 533–545.
- PARFIT, D. (1982) "Future Generations: Further Problems." *Philosophy and Public Affairs* 11(2): 113–172.
- PARFIT, D. (1983) "Rationality and Time." *Proceedings of the Aristotelian Society* 84: 47–82.
- PARFIT, D. (1984) *Reasons and Persons*. Oxford: Clarendon Press.
- PARFIT, D. (1995) "The Unimportance of Identity." In H. Harris (ed.), *Identity: Essays Based on Herbert Spencer Lectures Given in the University of Oxford*, pp. 13–45. Oxford: Clarendon Press.
- PARFIT, D. (1997a) "Equality and Priority." *Ratio* 10(3): 202–221.
- PARFIT, D. (1997b) "Reasons and Motivation." *Aristotelian Society Supplementary Volume* 71(1): 99–130.
- PARFIT, D. (2001) "Rationality and Reasons." In D. Egonsson et al. (eds), *Exploring Practical Philosophy: From Action to Values*, pp. 17–39. Aldershot: Ashgate.
- PARFIT, D. (2003) "Justifiability to Each Person." *Ratio* 16(4): 368–390.
- PARFIT, D. (2004a) "Overpopulation and the Quality of Life." In J. Ryberg and T. Tännsjö (eds), *The Repugnant Conclusion*, pp. 7–22. Kluwer Academic Publishers.
- PARFIT, D. (2004b) "What We Could Rationally Will." Tanner Lectures, 2002. In G. Peterson (ed.), *The Tanner Lectures on Human Values*, Vol. 4, pp. 285–369. Salt Lake City: University of Utah.
- PARFIT, D. (2006) "Normativity." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 1, pp. 325–380. Oxford: Clarendon Press.
- PARFIT, D. (2011) *On What Matters*, Vol. 1. Oxford: Oxford University Press.
- PARFIT, D. (2011) *On What Matters*, Vol. 2. Oxford: Oxford University Press.
- PARFIT, D. (2012) "Another Defence of the Priority View." *Utilitas* 24(3): 399–440.
- SIDGWICK, H. (1981 [1907]) *The Methods of Ethics* (7th edition). Indianapolis: Hackett.
- SINGER, P. (ed.) (forthcoming) *Does Anything Really Matter? Parfit on Objectivity*. Oxford: Oxford University Press.