INTRODUCTION TO THE COLLECTION

Andrea Sauchelli

This volume is divided into two parts: the first introduces Derek Parfit's *Reasons and Persons* (henceforth 'R&P'), whereas the second includes chapters that critically discuss recurring ideas in R&P. The chapters in this collection were written by different authors, and their styles and approaches slightly differ from each other. As the editor of this volume, I decided against imposing any strict requirements on its contributors, with the exception of reminding the contributors to the first part that their chapters are supposed to help the readers better understand the content of Parfit's book. Some of these writers adopted a more critical style, whereas others chose a more illustrative and exegetical approach. I think that they have all achieved the aim of introducing Parfit's book clearly, albeit in different ways. The chapters in the second part were commissioned with the intent of collecting works in various fields of philosophy that further elaborate on some of R&P's principal themes and ideas. As will emerge from this brief introduction, the variety of the areas of research discussed in R&P is remarkable.

Parfit's book has become a contemporary classic, widely read both by philosophers and scholars in other fields (e.g. psychology and even economics). Parfit made several changes to the first edition of R&P published in 1984—the introduction to the 1987 edition contains a brief summary of these alterations.[1] In its 1987 version, R&P comprises four parts and ten appendices. Regarding its content, R&P elaborates on several works that Parfit published from the early 1970s to the beginning of the 1980s. In fact, entire chapters are based on earlier material, albeit modified in light of the criticisms and suggestions Parfit received from an astonishing number of other influential philosophers (the long list includes the likes of Amartya Sen, Shelley Kagan, Larry Temkin, Bernard Williams and John Broome). Among the authors whose published works have more conspicuously influenced R&P, whether directly or indirectly, we may list: Henry Sidgwick, Thomas Nagel, David Wiggins and Bernard Williams. The success and enduring popularity of R&P

may be partially explained by its impressively high standard of argumentative rigour, the inventiveness and perspicacity of the case studies discussed, and the interesting and controversial theses defended throughout the book. Furthermore, R&P touches on a variety of different topics in apparently disparate fields of philosophy—among others, rational choice theory, normative ethics, personal identity and population ethics—and brings together various problems and issues that others (with rare exceptions) have discussed only separately.

Although R&P 'contains multitudes', there are recurring themes and unifying threads that run through it; for example, Parfit's attempt to show that one popular version of what he calls the Self-interest theory (S) is false.[2] To a first approximation, this theory about individual rationality tells us that each person has a supreme rational ultimate aim, namely, that her life go for her as well as possible. Because there are different conceptions of how a life can go well, and Parfit aims to provide arguments sufficiently general to apply to several versions of S, he painstakingly explores the applicability of his arguments to the various ways in which S can be further understood.[3] In turn, the recurring criticism of S is developed 'from different fronts'. More specifically, in the first part of R&P, *Self-Defeating Theories*, Parfit suggests that S, along with consequentialist theories of morality (C), may be indirectly self-defeating and possibly self-effacing. A theory T is *directly* individually self-defeating when there are cases in which it is certain that, if someone successfully follows T, she will thereby cause her own T-given aims (the aims given to her by the theory itself) to be more poorly achieved than they would have been if she had not successfully followed T. Parfit argues that S is not directly but *indirectly* self-defeating because there are people for whom it would be worse if they were disposed never to do what they *believe* would be worse for them. This point does not show that S fails on its own terms because, as a theory of individual rationality, S does not claim that each individual should never act irrationally. In fact, Parfit suggests that in certain cases it may be rational to act irrationally (there can be cases of rational irrationality)—perhaps just for a short time—*and* this is compatible with S. Although not directly self-defeating, S and C are self-effacing because they may both imply that we should try to believe in some other theory. For instance, C implies that we should believe the theory such that, if believed, the outcome would be best—and, crucially, this point is compatible with *not* believing C itself (a similar reasoning applies to S). In addition, and perhaps even more importantly, S can be *collectively* self-defeating. In particular, consider situations involving more than one individual in which: (i) the achievement of each person's T-given aims partly depends on what others do, (ii) what each person does will not determine what the others do, and (iii) T is agent-relative; that is, it gives to different agents different aims. Parfit claims that there are many cases in which if each person rather than no-one does what will be better for herself, or her family, or those she loves, there will be a worse outcome for everyone. Collectively, we would be better off if not everybody acted in a self-interested way. However, because S is a theory of individual rationality, we may argue that such cases do not prove that it is decisively refuted. The above criticism (being collectively self-defeating) also applies to what

is termed common-sense morality—the reason being that this form of morality is generally regarded as including the idea that we have special moral obligations towards members of our family, and that following such obligations may bring about situations that are collectively worse than those in which these obligations are not followed. As highlighted by Ben Eggleston in his introductory chapter, the discussion in Part One is functional to the outlining of the general traits of a new moral theory that does not suffer from the above problems and that unifies consequentialism and certain aspects of common-sense morality (see Eggleston's contribution for more details).

After having discussed in Part One some arguments that do not seem to directly refute S, in Part Two (*Rationality and Time*), Parfit proposes other arguments that are supposed to be sufficient to reject it. This section of R&P does not question our non-reductionist intuitions about our nature and continuity over time (more on non-reductionism later) and attempts to prove that S should be rejected for reasons that are compatible with different theories of personal identity. As outlined by Brian Hedden in his contribution, in this part of R&P Parfit offers three main arguments against S. In particular, Parfit suggests that it may be rational not to care most about one's own well-being and to care at least as much for other things, the pursuit of which we may believe is not conducive to the best possible outcome for ourselves. Examples of desires for these things include the desire to sacrifice oneself (or, at least, not to maximise our well-being) for moral reasons or desires for achievements (or, better, for some achievements in certain circumstances). The latter are specified in a vaguely Nietzschean fashion because Parfit includes among them the desire to produce a great work of art despite regarding the fulfilment of such a desire as not leading to what is best for oneself (within reasonable limits). Parfit's point is that these desires may be no less rational than the desire for what the relevant agent deems best for herself. The second line of reasoning against S is focused on one of its alleged faulty structural features, namely, the fact that such a theory is agent-relative (in specifying the aim that is rational for an agent to pursue, the theory makes essential reference to the agent herself) but time-neutral (in considering what is best for an agent, said agent should count the well-being of each temporal part of her life equally). Against this general structure, Parfit suggests that there are reasons to prefer a theory that is either fully neutral or fully relative. The third argument against S is based on the idea that it may not be irrational to be time-biased—for instance, to care more about some future parts of our lives rather than those parts in the past. As for the rest of R&P, the subtle thought experiments and ingenious reasoning used to argue for these points have been highly influential and have helped to shape the contemporary debate—for more details, see Hedden's chapter.

The third part of R&P (*Personal Identity*) contains another important and recurring theme of the book: the idea that changing our beliefs about our nature and persistence may have important consequences for various issues in moral theory and applied ethics. More specifically, Parfit's achievements in this part are at least twofold: first, he clearly delineates and forcefully defends a version of what he will later call *Constitutive Reductionism* (a family of theories of personal identity)

and, second, he investigates in some detail the practical and moral consequences of adopting such a view.[4] In this respect, Part Three decisively contributes to the debate on whether the holding of the relation of personal identity is a necessary component of the what-matters relation (R). In this context, R can be understood as the relation that determines the extension of our rational self-concern: when P at t_1 is R-connected to Q at t_2, P's well-being is part of Q's. One of Parfit's most debated theses is that, contrary to the opinion of many other philosophers, personal identity is not what matters. In this paragraph, I will briefly summarise only part of his lengthy reasoning for this conclusion. Parfit thinks that relation R is exhaustively composed of two more fundamental relations that only partly compose personal identity, namely psychological continuity and psychological connectedness when they hold in the right way.[5] The amount and relevance of direct psychological connections between two persons at different times, P and Q, determine the degree of psychological connectedness between P and Q. In several versions of the psychological view, when a strong degree of psychological connectedness is established, and chains of such connections hold between P and Q, we can say that P and Q are psychologically continuous. Examples of direct psychological connections between P at t_1 and Q at t_2 include P's experiencing of an event at t_1 *and* Q's recollection of it at t_2, Q's acting at t_2 out of P's intention at t_1, and so on. Crucially, given the nature of the relevant grounding relations, personal identity may be a matter of degree (this view contrasts with the theory that personal identity depends on a non-physical and non-psychological entity that is always determinate [e.g. a Cartesian Ego]). Due to the fact that personal identity includes a non-branching condition—roughly speaking, the relevant psychological relations should hold between at most two persons each at different times—and only the proper holding of the relevant psychological relations matter, personal identity is not a necessary condition for what matters. In short, personal identity includes psychological connections, continuity, the non-branching condition and, on some versions of this criterion, a condition regarding how these relations are supposed to hold (e.g. R should hold in virtue of the continuity of parts of the relevant person's brain). However, R does not necessarily include the non-branching condition. Therefore, there are cases in which personal identity and R do not coextensively hold. Parfit argues for this conclusion by elaborating on a thought experiment previously discussed by Sydney Shoemaker and David Wiggins. The upshot is that there are cases of symmetric fissions—cases in which an individual's relevant psychological connections existing at t_1 are equally distributed between two different persons each existing at a later time t_2—the outcomes of which it may be irrational to regard as being as bad as death. In this part of R&P, Parfit also elaborates on the consequences of adopting a reductionist view (the psychological account of personal identity delineated above is one form of reductionism) for other issues in moral theory. For example, he explores the idea that we may assign a different weight and scope to certain principles of distribution (e.g. equality) proportionally to the degree of psychological connectedness holding intra- or inter-personally. See my introductory chapter for more details.

Future Generations, the fourth part of R&P, begins with the claim that it is of utmost importance that a moral theory should address how we ought to behave towards future generations. In particular, an acceptable unified moral theory (perhaps of the kind outlined in Part One) should solve a series of puzzles and problems addressing, among other things, harm and beneficence towards future people. Some of these problems partly derive from the fact that our present choices affect not only the number and quality of life of future people but also their identity. For instance, the famous *non-identity problem* stems from an attempt to reconcile apparently plausible principles, some of which involve the existence of future people. In particular, some philosophers claim that an act can be wrong only if it makes things worse for some existing or future people (*bad* must be *bad for* someone), and that an act is not bad for someone if the act brings about the existence of such a person, provided that the life of this person is at least worth living (or, at least, existence-conferring acts, acts unavoidable for the existence of an individual, do not make the existence they bring about worse). Now, Parfit puts forward some cases involving actions that we would intuitively judge to be wrong but that are simultaneously unavoidable for generating lives that are at least worth living. For example, take the case of a 14-year-old girl who decides to have a child and whose socioeconomic situation clearly suggests that she is unable to provide her child with a good start in life. Had she waited for several more years, she would have been able to give a better start in life to the other child she would have had. The life of the child she gives birth to is worth living but significantly worse than the life she could have given to the *other* child she would have had if she had waited several more years. Many people agree that the girl should have waited but can we say that, by not waiting, she has thereby harmed her actual child? How can we explain our initial intuition that the girl should have waited? According to Parfit, a satisfying moral theory should solve this problem and meet other requirements. These requirements include: (1) Avoiding the *Repugnant conclusion*—roughly speaking, the conclusion that it is better to have a large population of people whose lives are barely worth living than a population of significantly fewer people but with a much higher quality of life; (2) Avoiding the *Absurd conclusion*—consider two scenarios: in the first, there is a huge population at t_1 with a quality of life higher than our planet now in which one person in 10 billion has a life of uncompensated suffering, whereas in the second scenario, there is a collection of populations of 10 billion each (as before, one person in 10 billion has a miserable life) that do not interact with each other (e.g. each group of 10 billion of these people lives at times after t_1). If we impose a local limit on the value of positive quantity but not on negative quantity (for example, if we believe that there is a limit to the positive value that an increase in quantity can have at a specific time but also think that the disvalue of an increase in uncompensated suffering has no upper limit), then the first scenario is bad (because the quantity of suffering is not outweighed by the increase of quantity of positive value), whereas the second scenario is good (because the increase of quality outweighs the quantity of uncompensated suffering). However, this asymmetrical evaluation is absurd. After a painstaking discussion of these problems and possible solutions, Parfit claims that,

in R&P, he has found no theory that satisfies all these requirements. See Roberts's chapter on Part Four for a critical discussion of these and related issues.

In one of the final sections of R&P, Parfit selects this general point as a common theme or lesson to be learnt from his book: 'our reasons for acting should be more impersonal'. As it has partially emerged in the brief introduction above, this idea has taken different forms—for example, the application of his reductionism in personal identity to morality, and his rejection of person-affecting principles to solve the non-identity problem. In a way, the rest of Parfit's career can be seen as an increasingly enriched attempt (mostly at a rather theoretical level) to further refine and improve on the conclusions reached in R&P.

The second part of this collection comprises new original papers on some of the ideas in R&P.[6] In particular, Chrisoula Andreou's chapter discusses some theoretical consequences of Parfit's quandaries (and later elaborations by other philosophers) on puzzles and problems in value theory. In particular, Andreou considers the transitivity of the "better than" relation, using Parfit's work on the *Repugnant Conclusion* as her starting point. Andreou considers the possibility of betterness cycles and the implications of accepting the intransitivity of "better than." She argues that if betterness cycles are indeed possible, then a distinctive form of satisficing that involves reasoning in terms of leagues, plays a crucial role in proper reasoning about what to do.

David Braddon-Mitchell and Kristie Miller's contribution outlines the conceptual terrain of what they call *conative* accounts of personal identity. These views have in common the idea that personal identity over time depends on conative phenomena such as desires, behaviours and conventions. In particular, the authors distinguish these conative views along three dimensions, namely, on the basis of (1) what *role* the conations play, (2) what *kinds* of conations play that role and (3) whether the conations that play that role are *public* or *private*. Braddon-Mitchell and Miller also evaluate such theories by adopting two key desiderata: accommodating faultless disagreement and accommodating our practical concerns.

Christian Coseru addresses the following questions: What justifies holding the person that we are today morally responsible for something we did a year ago? Further, why are we justified in showing prudential concern for the future welfare of the person we will be a year from now? Coseru suggests that these questions cannot be systematically pursued without addressing the problem of personal identity. His chapter considers whether Buddhist Reductionism, a philosophical project grounded in the idea that persons can be reduced to a set of bodily, sensory, perceptual, dispositional and conscious elements, provides support for Parfit's psychological criterion for personal identity. Coseru examines the role that self-consciousness plays in mediating both self-concern and concern for others, offering an argument for how reductionism about substantive or enduring selves may be reconciled with the seemingly irreducible character of self-consciousness.

Nilanjan Das and L.A. Paul investigate some philosophical aspects of a subclass of acts, namely, those acts that change who we are (*personally transformative acts*). A personally transformative act is one that brings into existence a future self that is

radically different from who the agent previously was. In some of these cases, the agent may be antecedently certain that the existence of this future self, although worth having, will be unavoidably flawed, even if the future self values its existence. However, if the agent does not perform the transformative act, she will not change so radically, so her unchanged future self may indeed be better off than her transformed future self. In their chapter, Das and Paul argue that situations of this kind raise a problem that is structurally similar to the non-identity problem.

In his contribution to this collection, Dale Dorsey unravels some important theoretical issues related to the Self-interest theory (or prudence). In particular, he discusses a problem associated with the idea that, although the Self-interest theory is not the whole story about practical rationality, many philosophers find it entirely plausible to hold that prudence is the best theory of rationality when it comes to normative self-concern, the idea being that, when our decision concerns only us, we have the strongest reason to promote our welfare to the greatest extent. However, prudence can seem alienating, especially in cases in which we are called upon to abandon deeply valued projects for the sake of projects we may have already taken on (or have yet to take on)—and yet, prudence seems precisely correct in cases of other, less significant welfare goods. Dorsey argues that this puzzle can be solved by holding that self-concern is not prudential. In particular, he claims that self-concern is not (or need not be) welfarist in nature.

Carol Rovane focuses her attention on Christine Korsgaard's early critical response to Parfit's *Reasons and Persons*, in which Korsgaard pointed out that Parfit's reductionist account of personal identity did not take due account of the fact that persons are agents. In her contribution, Rovane offers a reductionist account of personal agency that takes this into account. In particular, Rovane's reductionism holds that the existence of a person consists in nothing but a certain sort of intentional activity that stands in the right sorts of relations. The account also claims that persons are self-constituting in much the way that Korsgaard suggests. Rovane's form of reductionism, however, does not support Korsgaard's Kantian ambition to derive and ground an unconditional imperative of morality. Nor does it support the Kantian conception of the person of an end in itself, for it entails that persons, qua agents, exist for the sake of the ends that their existence makes it possible to pursue—the ends for the sake of which they constitute themselves. Rovane's account agrees with Parfit's claim that we must revise our common-sense notions about the moral significance of the individual person. Yet it does not invite the consequentialist orientation that Parfit thought his own reduction invited.

The last chapter of the collection, David Velleman's 'Non-identical and impersonal', discusses several topics through the lens of a broadly Kantian approach to ethics. In particular, Velleman offers a solution to the non-identity problem that resorts to the Categorical Imperative, thus rejecting some of the utilitarian assumptions that have characterised the debate so far. Velleman claims that rather than focusing on the notions of harm and benefit towards particular people, we should consider the idea that personhood itself can be disrespected. Velleman's chapter also contains a criticism of Parfit's theses on what matters—a criticism that

Velleman advances from the perspective of his imaginability-based account of what matters and personal identity.

Notes

1 In particular, Parfit (1984/87: x).
2 Even a lengthier summary of the book would be inadequate for capturing the richness of R&P. The brief introduction in the main text will sidestep many important issues and be imprecise in certain important aspects.
3 Parfit discusses in Appendix I, thoroughly analysed in Chris Heathwood's contribution, various different theories of well-being or welfare.
4 See Parfit (1999).
5 A more precise formulation of relation R is given in a later essay, that is, Parfit (2007).
6 With the exception of the short introduction to Velleman's chapter, the descriptions of the chapters in Part II in the main text are abridged versions of the abstracts sent by the authors.

References

Parfit, D. 1984/87. *Reasons and Persons*. Oxford: Clarendon Press.
Parfit, D. 1999. Experiences, Subjects, and Conceptual Schemes. *Philosophical Topics* 26, 1–2: 217–270.
Parfit, D. 2007. Is Personal Identity What Matters? *Marc Sanders Foundation*, 31 December. Retrieved at: www.marcsandersfoundation.org/wp-content/uploads/paper-Derek-Parfit.pdf

Additional resources

Dancy. J. ed. 1997. *Reading Parfit*. Oxford: Wiley.
A volume of important critical essays specifically focused on R&P. Parfit's replies are not collected in the same volume, and some have only been published online (e.g., Parfit, 2007).
MacFarquhar, L. 2011. How To Be Good. *New Yorker*, 29 August. Accessed 26 June 2019. www.newyorker.com/magazine/2011/09/05/how-to-be-good
An interesting profile that includes details on Parfit's life.
Rabinowicz, W. 2016. Derek Parfit's Contributions to Philosophy. *Theoria* 82: 104–109.
A concise and helpful summary of R&P and of some of Parfit's other achievements.
Symposium on Derek Parfit's Reasons and Persons, *Ethics* 96, 4.
Parfit modified the 1987 edition of R&P also in light of some of the comments and criticisms contained in this special edition of *Ethics*.
Williams, B. 1984. Personal Identity. *London Review of Book* 6, 10: 14–15.
Bernard Williams' review of R&P.

Parfit's publications prior and relevant to Reasons and Persons

Parfit, D. 1971. Personal Identity. *Philosophical Review* 80, 1: 3–27.
Parfit, D. 1972. On 'The Importance of Self-Identity'. *Journal of Philosophy* 68, 20: 683–690.

Parfit, D. 1973. Later Selves and Moral Principles. In A. Montefiore, ed., *Philosophy and Personal Relations*, 137–169. Routledge.
Parfit, D. 1976a. Lewis, Perry, and What Matters. In A. Rorty, ed., *The Identities of Persons*, 91–107. University of California Press.
Parfit, D. 1976b. Rights, Interests, and Possible People. In S. Gorovitz et al., eds., *Moral Problems in Medicine*, 369–375. Prentice-Hall.
Parfit, D. 1976c. On Doing the Best for Our Children. In M. D. Bayles, ed., *Ethics and Population*, 100–115. Schenkman Pub. Co.
Parfit, D. 1978. Innumerate Ethics. *Philosophy and Public Affairs* 7, 4: 285–301.
Parfit, D. 1979a. Is Common-Sense Morality Self-Defeating? *Journal of Philosophy* 76, 10: 533–545.
Parfit, D. 1979b. Prudence, Morality, and the Prisoner's Dilemma. *Proceedings of the British Academy* 65: 539–564.
Parfit, D. 1980. An Attack on the Social Discount Rate. *Philosophy & Public Policy Quarterly* 1, 1: 8–11.
Parfit, D. 1982a. Personal Identity and Rationality. *Synthese* 53: 227–241.
Parfit, D. 1982b. Future Generations: Further Problems. *Philosophy and Public Affairs* 11, 2.
Parfit, D. 1983a. Energy Policy and the Further Future: The Social Discount Rate. In D. MacLean & P. G. Brown, eds., *Energy and the Future*, 31–37. Rowman and Littlefield.
Parfit, D. 1983b. Energy Policy and the Further Future: The Identity Problem. In D. MacLean & P. G. Brown, eds., *Energy and the Future*, 166–179. Rowman and Littlefield.
Parfit, D. 1984. Rationality and Time. *Proceedings of the Aristotelian Society* 84, 1: 47–82.