

---

## *Reaction & Debate*

---

### *In Memoriam Derek Parfit (1942-2017)*

J. A. Colen – University of Navarra (Spain) and Minho University (Portugal)

#### I

Derek Antony Parfit's death occurred a little more than one year ago on January 1<sup>st</sup> 2017. His theories have stayed with us. Several major books, authored by both him and many of his contenders (and sometimes friends), were published in the very same year of his death, almost at the same time as large numbers of lengthy and deeper than usual obituaries came to print (Edmonds 2017; McMahan 2017; O'Grady 2017, etc.). Among the many books and essays on Parfit's moral theories, the volumes edited by Jonathan Dancy (1997), Simon Kirchin (2017), Peter Singer (2017), together with the volume authored and edited by Parfit himself (2017b), are among those that go deeper into the problem that Parfit left unresolved. This unresolved problem concerning his peculiar engagement in meta-ethics, is not, as will be argued here, the problem of the practical applications of the theory. The real problem is that he never answers –and tries hard to avoid answering – the question of how we can understand, respond and eventually act driven by (non-natural) reasons (2011, 31). Before articulating the problem more fully, we are in profound need of a map to the largely unexplored field of moral inquiry set by Parfit.

This mass of recent literature may help us to draw a provisional balance of his contributions to moral philosophy, which is unfortunately left incomplete by his death. He promised a fourth volume, which he will no longer be able to write, but whose content is not difficult to guess. It would have pursued the broad lines defined in the final pages of his last book: “One thing that greatly matters is the failure of we rich people to prevent, as we so easily could, much of the suffering and many of the early deaths of the poorest people in the world [...]. What now matters most is how we respond to the various risks to the survival of humanity” (2017b, 436).

After a very brief biographical note, I will start by describing the path to progress in moral theory that Parfit follows in his earlier work, comparing his method with those

of some of his ‘nemeses’, Bernard Williams and Robert Nozick, in order to clarify why the problem of our response to moral reasons cannot be solved in Parfit’s own terms. I will try to make it explicit why he turned to meta-ethical inquiry at the core of his triple theory and, to conclude, I will point out the most important questions discussed in the works now made available. Scholars in recent moral literature express major reservations about – or entirely dismiss – Parfit’s claim that there are objective normative truths irreducible to the natural world. This text explores the application of his peculiar method of ethics and articulates the question that is at the root of this prevailing scepticism. At the root of the unsolved problem is his peculiar method of ethics, which rests on moral intuitions that may simply be ‘revised’ or refined common sense, made more consistent – that is, unless some explanatory problems that are ‘genuine and real’ are addressed.

## II

Derek Parfit was born on December 11<sup>th</sup> 1942 in Western China, where his parents practiced medicine in Christian missionary hospitals. His parents moved to Oxford just one year after Parfit’s birth. He was at first religiously devout, but lost his faith when he was eight years old, disturbed by the problem of evil (if God is all-powerful, how can he allow evil to prosper in the world?). He was educated at Eton and majored in history at Balliol College, Oxford (1961-1964) and, after graduating, went for two years to Columbia and Harvard. In 1967, he gained a fellowship to All Souls. His first published paper, on “Personal Identity” (1971), granted him recognition. He wrote two books, and two books only, 17 years apart, *Reasons and Persons* (1984) and *On What Matters* (2011 – the third volume of which (2017b) has just been published by Oxford University Press six years after the first. At the end of his first book, *Reasons and Persons*, he asserts that “[...] the history of ethics may be just beginning” (1984, 453). Because “[...] few atheists made Ethics their life’s work” (1984, 454), nonreligious ethics is still a new and fertile field of inquiry. He reasserts this idea in his final volume: the field is still largely unexplored. His books circulated long before coming out in print form and were collaborative: not only did he engage in debates with his peers, but he included pieces they wrote and his (and their) own answers to them.

His books were also intensely personal and “Parfit himself also somehow seemed to live his theories, helped by perhaps having [...] Asperger syndrome” (O’Grady 2017). He first shied away from too abstract ‘meta-ethics’, which he deemed too difficult to address (2017b, ix; Temkin in Singer 2017, 2), demonstrating progress in moral theory in practice, but when the much expected first two volumes of *On What Matters* finally showed up, almost twenty years later, the bulky and somewhat disconnected book, the unity of the parts of which is far from obvious, proved to be largely dominated by meta-ethics. As he asserts, several theories are not just ethical but “[...] *meta-ethical* in the sense that they are about the meaning and truth of moral claims, and of other normative claims” (2017b, 2).

Which theories of morality share this meta-ethical ‘core’? In Part I of *Reasons and Persons* Parfit begins by distinguishing theories of morality (“we ought to try to act morally”) from theories of rationality (“we ought to try to act rationally”) in accordance with the formal goals of the theories (1984, 3). But there are also theories with different *substantive* goals, that is, that are not merely-formal. He discusses some of them to make the case that they are “self-defeating theories” and fail on their own terms. We therefore need to “revise” (2017b, ix) them or new ones should replace them. Among the theories that are self-defeating there is in particular the self-interest theory of rationality (S) and two other ethical frameworks: Common-Sense Morality (CSM) and Consequentialism (C).

Self-interest theories can take different shapes, such as Hedonistic (everyone seeks *his or her own* pleasure or happiness), Desire-Fulfilment (seeking not happiness but the satisfaction of one’s most intense goals, that give life its meaning despite potentially undermining happiness) and Objective List Theory (like most natural law theories, but also all theories that uphold that contemplation or a healthy life are parts of an objective list of what is good for oneself). These different theories partially overlap (2017b, 4). Parfit assumes that self-interest theories of some kind have been pervasive in the west until this century (2017b, 86; 194), equating rational self-interest with the aim of ensuring that our whole life goes ‘as well as possible’, disregarding sincere altruism. All self-interest theories hold in common that it is irrational to commit any acts of self-denial or to act on desires that negatively affect our wellbeing. One may consider a writer (Kate) whose strongest desire is to write a great novel but who, in doing so, suffers from exhaustion and depression (2017b, 6; 13-17). Parfit argues that we have such desires in conflict with our own wellbeing, and that it is not necessarily irrational to act to fulfil such desires.

Henry Sidgwick (1981/1874) longed for the fusion of ethics and rationality, while Parfit admits that acting irrationally and acting immorally are different things (see, for example 1984, 12-17; 22; etc.), so he does not uphold the union between the two. Theories of morality and theories of rationality just present formal aims. Aside from the initial appeal to the plausibility of desires that do not directly contribute to one’s life being the best, Parfit was able to find instances of situations in which self-interest is *indirectly* self-defeating – that is, its demands initially seem irrational: for example, demands to act like a lunatic if your house and family are being burgled and there is no good rational outcome (2017b, 12-17; see his answer to Gauthier in 2011, 433ff). Self-interest does not fail on its own terms (2017b, 11), but it seems to suggest the adoption of an alternative ethical framework, or perhaps a self-effacing version of it (2017b, 23). For instance, in a transparent world, it might be in my self-interest to become trustworthy in order to participate in mutually beneficial agreements, even if later in maintaining my agreements I will be doing what will be worst for me. One such instance deals with keeping promises to those who give you a car ride (2017b, 7). Jonathan Dancy thinks this kind of self-defeating situation is in effect tantamount to full-fledged self-refutation. This would be clearer if more persuasive examples were used to appeal to “[...] changeable empirical claims about human psychology” (Dancy 1997, 27; see 1984, section 63, chapter 8).

Parfit does however seek to avoid psychological and other ‘naturalist’ or deterministic explanations, resorting instead to the popular ‘prisoner dilemma’ and similar (old and newly invented by him) moral dilemmas such as the ‘Samaritan’s’ or ‘commuters’ dilemmas. The most common moral dilemmas are those in which many contributors are involved (2017b, 61) and some mistakes about “moral mathematics” are widespread, such as the idea that very small contributions do not matter (2017b, 67ff). Parfit contended that even if it was not indirectly self-defeating individually, both self-interest and consequentialism would be directly self-defeating collectively (1984, 27-28).

But while self-interest theories and consequentialism fail indirectly (2017b, 49-51), common-sense morality is collectively directly self-defeating. The same applies to self-interest, but self-interest is not a theory about collectives. Common-sense morality (2017b, 95) is not the popular morality of common sense or folk morality that upholds, say, that ‘we should not kill’, or that ‘it’s wrong to lie in court and condemn an innocent person’ (which may instead be instances of self-interest theory of the objective list sort).

‘Common-sense morality’, ‘self-interest’, like most of Parfit’s vocabulary, are terms of art and do not mean what most people would think they mean. The vocabulary is in part taken from Sidgwick’s *The Methods of Ethics*, “the best book on ethics ever written” (2011, xxxiii). Common-sense morality is, so to say, the idea that there is nothing wrong with caring for our family, friends and country first and foremost: “[...] most of us believe that there are certain people to whom we have special obligations” (1984, 95). Parfit showed, using interesting (but not always compelling) examples and often borrowing with a grain of salt from games theory, that it would often be better for us all if we *did not* put the welfare of our loved ones before all else (which is a mistake from anon-impartial-perspective). For example, we should care not only about our children, but everyone’s children and this will be the best for our children as well (2017b, 100; 103ff), even if “[g]reater impersonality may seem threatening” (1984, 443; see chapter 4 and Part 3).

Sidgwick’s book ends with a sad note – ‘failure’ was its last word (see Parfit 2017b, 443) – about successful conciliation of the dualism of reasons: agent-centred and universal benevolence – a conflict that he deems “the profoundest problem of ethics” (Sidgwick 1907, 336, note 4). Parfit is somewhat more optimistic, although less than Peter Singer thinks he should be (see Katarzyna de Lazari-Radek and Peter Singer in Singer 2017, 292-295; *sed contra* see Adams 1997).

Fighting moral sceptics (1984, 114; 452-453), parts II and III of *Reasons and Persons* set out to pursue this line of moral reasoning and tackle the question of which is the best moral theory left standing, if we consider variable ‘time’ as part of the theory. The goal is to show even better which theory is (directly or indirectly, individually or collectively) self-defeating if we compare present and future goals. Current theories of rationality are found wanting. Parfit presents his Critical Present Aim Theory (CP), a theory articulated to accommodate any other competing theory (2017b, 131), to dismiss self-interest as our overriding rational aim and to allow the time of action to become critical. Self-interest or consequentialism are just instances, molehills in a larger landscape. He begins already in 1984 to climb the mountain of moral theories. Part IV is

supposedly an application of the theory that “[...] covers how we affect future generations”, but is in fact its “most important part” (2017b, 351) since it not only shows how fertile this new field of non-religious ethics can be, but deals with the significant harm we impose on ourselves and others through “[...] pollution, over-farming, soil erosion, famine, and overpopulation” (2017b, 444). Parfit’s (largely) conceptual theories are supposed to show progress in moral theory by example. Despite the very abstract nature of his books, or at least large sections of them, Parfit influenced the debate on practical moral questions and public policies. With, in particular, his views on the ageing population problem and global warming he was able to reach a wider audience.

### III

What demonstrates how fertile Parfit’s theory can be? Following the path to progress in moral theory in his earlier work, we need to compare his method with those of some other contemporary moral theorists that he opposed, while addressing the similar question of personal identity: his ‘nemeses’ Robert Nozick and Bernard Williams. Along with Thomas Nagel, with whom he agrees more often than not (2017b, 3; etc.), they are the philosophers most often quoted in the book, and they represent the clearest alternative accounts of personhood.

We must therefore briefly examine some recent debates and try to understand the reason for the impasse that leads some philosophers to entirely dismiss the question of ‘personal identity’ or, like Parfit, stumble on problems and paradoxes related to persons’ identities. In this vein, Parfit debates the non-identity problem: if with different outcomes of diverse policies, there would be different people, why is any policy better than any other? (see 1984, 351ff); the repugnant and absurd conclusions (the larger the future population is, the larger the sum of happiness, however drab some lives become [2017b, 381ff], no life has worth if there is not a limit on quantitative negative life value [2017b, 391ff]); and the “mere addition paradox” (while counting future population happiness, it seems that what happens to the best-off matters more than what happens to the worst-off [2017b, 420ff]).

Let us take a step back. The identity of a person as a contingent individual specificity (say Kate) and identity as the permanent properties of an individual (or class of individuals, say human beings) are obviously related. But how? At a certain point in his first book, Parfit endorsed the Wide Psychological Criterion of connectedness or identity: if teleported to Mars, we do not need the same brain to be the same person, just psychological continuity (1984, 208; first edition only). He later claimed that we should not decide between criteria. In problematic cases, identity should remain indeterminate. Let us have a closer look at what is at stake. The concept of ‘person’ was not introduced into philosophy without a cost: the Greek word designated the mask used by actors in the tragic theatre (see Meyer 1993) and therefore Christian theologians had to clarify that by person they meant “an individual substance of a rational nature” (*S. Th.*, I, q. 29). ‘Person’ designated not only an outward manifestation, but an individual being. Similarly,

those who today think like Peter Singer or Roger Scruton that dolphins or pigs have intelligence justifiably call them “non-human persons” (see Singer 1985, 40-51; etc.). Christians theologians reserved the expression for Divine Persons, angels and men (see, for example, Plantinga, Thomson and Lundberg, 2010). Unlike the legal or ‘forensic’ definition of person, which is clear but largely arbitrary, when the question is considered in terms of the ‘nature’ of those who possess certain ‘mental’ or ‘moral’ characteristics (Rudder Baker 2000), the philosophical concept, at first glance, pertains to an individual whose nature is rational from cradle (and even before birth) to grave: idiots and infants alike (see Ryle 2009, 1 and Parfit 2008, 203). The issue becomes cloudier when we cease to think in terms of ‘natures’ and try to avoid metaphysics altogether – or at the very least a certain kind of metaphysics that posits the existence of non-natural beings or properties.

This seems an acquired taste, which in the public opinion was perhaps achieved less by sophisticated philosophical reasoning and more by literature and plays that discarded “dormitive properties” (Molière 1871). The opening phrase of Descartes’ *VI Meditation* states: “Nothing further now remains, but to inquire whether material things exist” (Descartes in Eaton 1927, 145). The question of the world’s existence remained as unanswered as ever in essays by Nozick and Parfit but, taking for granted the evidence of our rational intuitions, the *explanandum* is not if there is a material world, but given the material world, “why [there is] anything, why this” world (see Parfit 2011, 623-648; Nozick 1981, 115-164).

Descartes true concern was to show that what is outside the mind consists of pure geometric extension (see Gassendi in Eaton [1927, 245-246]; answer by Descartes in Eaton [1927, 262]). This is a direct consequence of Descartes’ method: what is true of the concept is true of the thing itself, so we must start by purging the qualities, nature, forms and other vestiges of Aristotelian metaphysics that dominated philosophy for centuries. If the efforts of some of Descartes’ followers such as La Mettrie led his theories to unexpected conclusions (La Mettrie 1996, 36) about humanity’s machinal nature, the despair of Descartes himself in finding, somewhere inside the brain – that is, in the infamous pineal gland – the connection between body and mind, introduced the brain for the first time to the field of ‘metaphysical’ concepts. Recently, according to Parfit, Thomas Nagel simply follows through and suggests that what we essentially are is our brains (Nagel 1986, 37-43; *sed contra* Palmer 2014; Parfit 1984, 273-274).

The brain is yet another unknown domain and uncharted territory whose map is still full of surprises. Since the 1960s, however, the analytical theory of personality has suffered a strange twist due to the appearance of a curious problem. Studies of the brain have shown the possibility of ‘fission’ between its two halves, left and right, raising the question: “Which of these halves am I?” (See Olson 1999, 46-51; 2003, 328-348). In the decades since, brain transplants have joined with brain fission, fourth dimension travel and other conceptual experiments in an attempt to elucidate how a ‘person’ is ‘divided’ or transformed over time. “Personal Identity Through Time” is the subject (and title) of the first chapter of Robert Nozick’s book *Philosophical Explanations* (1981, 29-70). In this text, Nozick reflects on a paradox, or rather a couple of puzzles that the English

philosopher Bernard Williams posited in a journal paper in 1970 entitled “The Self and the Future” (1970, 161-180) printed just before Parfit’s 1971 essay.

In a nutshell, the paradox that Williams presents is this: two individuals, A and B, see their memories transferred to each other’s bodies. Before the exchange, A and B are informed that one of the bodies will receive \$100,000 while the other will be tortured. Williams notes that regardless of what actually happens to the ‘A-body-person’ with B memories, this A person identifies himself with the earlier B person, and vice versa (Williams 1973, 47-48). This seems to indicate that concern for what happens to oneself in the future does not necessarily involve what happens to one’s body, challenging the “[...] philosophical arguments designed to show that bodily continuity [is] at least a necessary condition of personal identity” (2017b, 51), and it suggests that it is reasonable for people to identify themselves instead with their own memories, impressions, etc., and not with their body.

But Williams also features a second thought experiment in which A is only informed that he ‘himself’ in the future will be tortured, and that at that future time he will not remember anything that relates to his own past – his images and impressions will come from B. Fear of torture by A, despite the (anticipated) complete psychological dissociation, leads the author to conclude that the fear that, in spite of everything, plagues A is based on the conviction that “[...] my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time” (2017b, 53). The second case appears to indicate, unlike the first, that bodily continuity is integral to identity. An A-body-person faces ‘risk’ when deciding to transfer the prospect of torture to the B-body-person, and this is the risk that Williams considers – ‘perhaps neurotically’ says Williams (2017b, 59) – the essential feature of the personal identity problem. Each thought experience leads to opposite results, despite the situation described being objectively identical, and the impossibility of arriving at a logical conclusion adds to our perplexity. Williams, however, ends with an observation. There are aspects of personal identity related to bodily continuity and aspects of ‘mental’ continuity, i.e. experiences and memories (2017b, 62). In general, it is assumed that this dichotomy coincides with yet another dichotomy: aspects of personal identity in the first person (mental dimension) and third person (the body). The two examples he gives show an exactly inverse relationship (2017b, 64ff; esp. the objections 70-81).

Nozick chooses precisely this, among the many puzzles surrounding the issue of identity, to articulate the problem that also concerned Parfit’s first paper: “[...] how, given changes, *can* there be a continuous identity of something from one time to another, and in what does this identity consist?” (Nozick 1981, 29). Nozick’s purpose is to explain the assumptions underlying Williams’ paradox. What is implied in the paradox is essentially as follows: for something (x) to be the same as something else (y) through time, it is assumed that identity depends on certain properties of both (x, y) and their mutual relationship, but that no other factor should be considered in this context.

Is this so? Nozick gives a graphic example and explores its transposition to the case of personal identity: Vienna Circle survivors meeting again in Istanbul and considering themselves the followers of the same group, only to discover afterwards that the remaining

members took refuge in the US. Which is the ‘true’ Vienna circle? What defines identity over time? (Nozick 1981, 29-31). This metaphor sheds some light on the issue and leads him to think that what is called identity through time is the “closest continuer” (1981, 33). The notion to which we refer when we speak of the identity of things is *continuity*, which allows us to provide the framework for an answer, but not yet to fill in the details thereof, because there are at least two important properties to consider: spatial-temporal continuity and continuity of features or physical elements. Put differently, the situation resembles a Lockean puzzle, Theseus’ ship, the planks of which were replaced periodically, only for it to be discovered later that another boat was fully restored with the original boards. Continuity in space and time points to the Theseus’ ship that was kept going through gradual maintenance; but the continuity of the physical elements points to the ship that, in this example, is born again from the original boards that had been stored (1981, 29-33). What Nozick calls ‘applications’ of this continuity theory to the problem of personality are various cases in which the brain is duplicated, including memories (case 1); a transplant (case 2); brain patterns transferred from a dying person to a new person (3); only half of the brain transplanted (4); or removed (5); or both simultaneously (6); or even a similar brain generated at random in the infinite universe (7). Not only this: Nozick also notes that it is not impossible, as in the Vienna Circle analogy, for there to be an overlapping of two identical persons (1981, 37-40). These ‘thought experiments’ allow him to clarify what we mean by the continuity of personal identity over time and to establish the conceptual approach to tackling the problem. He suggests notions such as ‘relational’, ‘closest relative’ and more curiously ‘intrinsic abstract structures’ (1981, 47-48), a concept somewhat analogous to the Aristotelian ‘form’ mocked by Molière (see *Metaphysics* I with Nozick 1981, 47). The surprising conclusion is that even in most of these extreme situations we can predict in which cases a person will have an identity over time with the simple model of the ‘closest continuer’, while ignoring the underlying complexity. Underlying complexity is the short name for all the mental and physical differences, impressions and memories of each and every individual, or to put it briefly: the set of characteristics that define *a* person as *a* contingent instance of ‘personhood’ (1981, 110-114). Nozick, however, as a libertarian in both politics and metaphysics, defends free will and, unlike La Mettrie, believes that we are not mere ‘machine-men’. If Nozick kept the problem on the philosophical agenda, Derek Parfit led it to perfection from his first paper (1971) up to the book *Reasons and Persons* (1984, Part III). Science fiction stories such as Star Trek’s including tele-transportation through space and time entered his ‘non-metaphysically weighty’ claims.

It is obvious that some at this point will ask themselves (as Quine in fact did): “The method of science fiction has its uses in philosophy, but [...] I wonder whether the limits of the method are properly heeded. To seek what is ‘logically required’ for sameness of person under unprecedented circumstances is to suggest that words have some logical force beyond what our past needs have invested them with” (Quine 1972, 490; Parfit does not ignore the challenge. 1984; 1999).

Parfit, however, is convinced that these imaginary situations involve something more than words: deep beliefs about what personal identity is (1984, 179). He first describes the



space-time continuum, which is the normal identity criterion for physical objects, and then the psychological and physical criteria. His purpose is to clarify the false assumptions of ‘physicalism’ or materialism, that is, the thesis that there is nothing real about mental states, which are no more than another kind of physical event. Not all philosophers are materialists but, according to Parfit, those who are not ‘physicalists’ are dualistic or idealists. What he tries to show is that, paradoxically, materialists might accept the psychological criterion of continuity of the person and dualists might accept the physiological criteria of temporal continuity. This paradox seems revealing to him. As there is no adequate criterion of personal identity through time, he concludes that persons are not like a Cartesian Ego, but are “like nations” (1984, 275). An alternative account should consider the immunity to error identification as regards first person nouns (Husserl 1959; Wittgenstein 1958, 66-67; Shoemaker 1968).

Parfit argues that reality can be described impersonally (from Mars). There does not, therefore, have to be a definite answer to the question “will I continue to exist?”, even after examining all the data involved. It is a mistake to assume that what is of interest in personal identity is survival: what matters is the Relation R, psychological connectedness (especially that established by memory and character) or continuity “with the right kind of cause” (2017b, 262). The controversy is not exhausted, therefore, in the problem of the criterion of defining identity through time. There is also a debate as to whether personal identity matters, or personal survival, and some such as Samuel Scheffler suggest that the collective survival of humanity is more important (Scheffler 2013, 15ff). Parfit could not agree more (1984, 281ff). Our existence is not a fact that must be “all or nothing” (1984, 281). Thus, the separateness of persons that Sidgwick deemed a deep truth leads in fact to the illusion that “[...] there is no justified sacrifice of some of us for others” as Nozick thinks (1984, 330; 520, note 93; see appendix E; Nozick 1974, 33). On the contrary, because we are deeply connected, what happens to all the others, even in the far distant future, matters. Moreover, we are currently creating risks for the survival of humanity. What really matters is how we reduce these risks allowing our descendants to eventually spread “through this galaxy” (2017b, 436).

What is really at stake in these debates, involving more than a hundred titles since 1970 – not to mention the countless scientific papers – seems to be always centred around the need to clarify the concepts or the beliefs embedded in language. This is why Quine’s question seems so relevant: do words have some power beyond the force which the past social need for words invested in them? The risk is that *puzzles*, challenges and arguments show us no more than the logical consequences of our own convictions or opinions, even prejudices. Of course, sometimes the results are counter-intuitive – a situation that should not be surprising, since the task that science and philosophy have imposed on themselves for three centuries is to replace the concepts of common sense with more rigorous artificial ‘constructs’ (in the same vein, see Driver in Kirchin [2017, 172-187; esp. 181ff with Parfit’s answers [2017b, 232-234]).

The perplexity that remains after revisiting this controversy is whether or how philosophers know *much more* than their method allows. It is not clear how the analysis of language leads to Bernard Williams’ physicalism. Williams seems to know more than

his *aporias* may suggest and, in a subsequent essay, he refutes the (four) objections that still upset the affirmative answer to the question “Are people bodies?” (1981, 64-78). What supports Robert Nozick’s free will, despite being a ‘reductionist’ (see Parfit 1984, 477; Bratman 2007, 107ff) and releases Derek Parfit from his own ‘Self’ in a kind of diluted collective? He found this “liberation from the self” refreshing, not depressing (Parfit 1984, 281). Imprisoned in a glass tunnel, the walls of the tunnel suddenly disappeared.

But in each and every case the dichotomy between psychological (or mental) and physical properties remains and we are thus led to conclude that, whatever we think Parfit claims about the R-Relation or personhood connectedness, we remain imprisoned in Cartesian dualism, and not merely in Sidgwick’s dualism of reasons. Parfit took the metaphysical question seriously in *Reasons and Persons* but abandoned it later for “lack of evidence” (Parfit 2008, 206; endnote 2; 1984, 275; etc.). Without addressing it however (or attempting to do it), going further than prevailing ‘intuitions’ imbedded in linguistic practices, as revealed by ‘thought experiments’, he cannot find the resources to clarify why “[w]e are animals that understand and respond to reasons” (2011, 31) – or so we claim.

In fact, despite promising progress in moral theory, we are inclined to think that the question ‘why morality?’ cannot be solved either in Williams’, Nozick’s or in Parfit’s own terms. But Parfit was certainly engaged in seeking a consensual answer for it, while avoiding ‘weighty ontological implications’.

#### IV

We have found ourselves so far with a need to take the long road and go back to the first sections of *Reasons and Persons* (1984) and the older (2011) volumes of *On What Matters*, intent on assessing Parfit’s theories with more than the charitable tones of his eulogies, and less than the strait-jacket of ‘non-analytical’ naturalism or any other technical view in meta-ethics. His terminology may be ‘baffling’ (a recurrent word in the commentaries), although Parfit engages several times in clarification (see scheme in 2011, 6; 263ff). The preferred terminology changed slightly during the composition of Parfit’s work, but in the last volume of *On What Matters* he gives us again a rough survey of the field of meta-ethics as he sees it. Because Parfit’s goal was “[...] to rechart the territory of moral philosophy” (Samuel Scheffler in 2011, xix), we inevitably find ourselves in uncharted territory.

To sum up, the tree of moral theories separates into two branches. The first branch carries (i) ‘non-cognitivists’, who deny that any normative claims *intend* to state truths, and (ii) ‘nihilists’ or ‘error-theorists’, who believe that most claims are intended to state truths, but all are false. For all of these, moral knowledge is impossible and nothing really matters to the sceptics and nihilists. A second branch carries all (iii) the cognitivists who believe otherwise. Among these, some are ‘normative naturalists’ who think that morality and normative claims are “like other truths about the natural world which might be empirically

discovered,” but some others are ‘non-naturalists’, according to whom “some normative claims state irreducible normative truths,” that are not mere “causal or psychological facts” about the best means to achieve any of our aims. For all non-naturalists, including Parfit, moral claims do not correspond to how things are but are moral *intuitions*, similar to logical, mathematical, or modal claims. Moral truths are like truths about triangles that apply even if there were no instances of triangular things. “And when some acts are right, or wrong, these moral truths could not be the same as certain natural facts, such as the fact that these acts would, or would not, minimize suffering, or would be acts of which most people would approve or disapprove.” (2017b, 2-4; see chapters 38; 39).

In *Reasons and Persons*, Parfit raises questions about which actions are right or wrong but shies away from meta-ethics. He thinks that the best way to uphold objective morality and progress in moral theory is to show that it is fertile. When writing, Parfit is often lively but highly convoluted, such that he asserts in his own introduction to the book that “many introductions [...] try to explain the central concepts that are used. Since it would take at least a book to give a helpful explanation, I shall waste no time in doing less than this” (1984, ix). With *On What matters*, by contrast, the first two volumes begin with a hefty 58 page introduction, to which a 37 page set of summaries was then added in the third volume, which may indeed provide a sense of the vast scope of topics that was found wanting in *Reasons and Persons*. But despite Parfit’s efforts to justify the order of the volumes (2017b, 435), these lack any obvious unity. No real attempt is made to justify the inclusion of all the sections and appendices in the book. Samuel Scheffler’s introduction focuses (2011, xixff) on Parfit’s Tanner Lectures, on Kant’s contribution to moral theory (in parts 2 and 3), the criticisms of the Tanner lectures by his peers and Parfit’s own responses (parts 4 and 5). He points out that part 1 is “[...] an extended discussion of reasons and rationality,” providing the background for the Tanner Lectures, and part 6 takes up “meta-normative questions raised by our use of normative language” (2011, xix).

As Mark Schroeder noted in a critical review, the volumes are really a set of distinctive monographs, including the monograph on rationality (section 1), and an “immense monograph on meta-ethics” (Schroeder 2011). The first volume of *On What Matters* begins famously with “Climbing the Mountain,” a work of normative ethics aimed at showing how three seemingly diverse approaches to normative questions (rule consequentialism, Kantian normative ethics, and contractualism) were reaching the same normative truths from different starting points. But Parfit decided he also needed to defend meta-ethical ‘non-naturalism’ as part of this project, so he added a second volume.

As Kirchin observes, Parfit’s attitude concerning meta-ethics in volume II contrasts with the conciliatory efforts towards Kant, Scanlon, and rule consequentialism in volume I. He “[...] criticizes, sometimes quite sharply, several prominent contemporary philosophers” (Singer 2017, xi). Indeed, some are said to be mistaken or ‘deeply misguided’. Even in his more recent (2017) responses to the essays included in the volume Kirchin edited, *Reading Parfit*, we find expressions of impatience (and some amicable arrogance) such as these: “These remarks puzzle me” (Parfit in Kirchin 2017, 189); “[...] in these pages I believe that I go further than most other philosophers in claiming that...” (2017, 190), etc. Despite some compelling arguments behind the Triple Theory, proponents of

‘naturalism’, ‘non-cognitivism’, and ‘subjectivism’ were not climbing the same mountain, or so it seemed at the time of publication of the first two volumes in 2011.

Perhaps the ‘ecumenical’ approach to the diversity of moral theories found in *Reasons and Persons*, and in the first volume of *On What Matters* had its limits after all. It did not extend to the rival meta-ethical views discussed in volume II, and perhaps they could not be further extended except at the cost of ironing out some rough edges. Kirchin’s book deals with volumes II and III only (a wise decision, considering Parfit’s publication schedule, which postponed Dancy’s collection for 13 years). But in Volume III, the attitude changed again after Peter Singer’s book. Singer shared with Parfit the idea that morality is objective and edited a volume of essays on meta-ethics, entitled *Does Anything Really Matter? Essays on Parfit on Objectivity*. Parfit found it disturbing that people like Bernard Williams (or Peter Railton, Sharon Street, Alan Gibbard – and Friedrich Nietzsche!) defended views in such frontal disagreement with his own. If his arguments were really as powerful as he thought, could not perhaps these brilliant theorists be brought into the fold? Overall, volume III is precisely an attempt to show that also in the domain of meta-ethics there is less disagreement than meets the eye, though bringing Nietzsche to the fold is perhaps going too far and at most what can be done is to ‘disarm’ Nietzsche’s challenges to morality (Andrew Huddleston in Singer 2017, 170). Disagreements are supposedly often misunderstandings, and therefore residual; some will eventually almost disappear in the best versions of some of their brilliant theories. Gibbard and Railton warm up to the idea (2017b, 8; 9; 10-11; 12).

In volume III, Parfit makes an attempt at ironing out some rough edges, but this comes at a price, namely diluting fundamental ethical alternatives. Nozick is – despite his own claims – a reductivist; he presents Kantianism in a way that would have made Kant faint or perhaps awake yet again from dogmatic slumber. The best theory left standing is nonetheless some version of Parfit’s own, *non-realist* cognitivism. But, as he explains, he is not committed to any kind of metaphysical view of this universe: “Non-Naturalist views can differ ontologically by making different claims about what exists and what is real [...] *Non-Metaphysical* Non-Naturalists make no such claims, since these people deny that irreducible normative truths have any such ontologically weighty implications. Nagel, Scanlon [and Parfit himself] accept one such view” (2017, 4). Metaphysics has a bad name, but most of his colleagues find Parfit’s ‘irreducible normative properties’ just as mysterious as any metaphysical claims.

Perhaps Parfit’s attempts at ‘correcting’ his closest peers, even if deliberate, was misguided or merely guided by the aim to perfect his own theory, sometimes at the cost of re-interpreting his fellow philosophers’ thoughts in a way that they would not have accepted, but the rationale behind his concern with moral language and meta-ethics, is this: he came to believe that most disagreements stemmed from a lack of – or misunderstanding of – the basic moral categories. As Mark Schroeder noted a few years ago, with some impatience: “I’m flattered to report that [unlike Bernard Williams among others], I am among the few meta-ethicists, whom Parfit credits as sharing the required conceptual repertoire to disagree with him. But unfortunately for me, it turns out that we don’t disagree either, because, according to Parfit, I don’t believe my own theory” (Schroeder 2011).

## V

What questions are discussed in the literature now made available? The contributors to Simon Kirchin's and Peter Singer's recent volumes, like several others before them, seem to think that Parfit's theory needs as much fine tuning as he deemed necessary to apply to salvage Kantian theory, Scanlon's contractualism and rule consequentialism in order to bring them together.

The very first paper in Singer's volume raises this question, which appears indeed appropriate for drawing a balance, if seemingly harsh: "Has Parfit's Life Been Wasted?" The author, Larry Temkin, also the author of *Rethinking the Good Moral Ideals* (2012), argues that Parfit did not waste his life, not because he was right, but because his claim that his life only mattered if he was right is too extreme. As Parfit states, either some version of moral realist externalism, such as non-naturalism (similar to his own), is true, or something that is not very different from nihilism is true. The strength of his conviction was so powerful that he declared his life (and that of some other moral philosophers such as Sidgwick and Ross) *almost* worthless if he was mistaken (Temkin in Parfit 2011, 367). The essay kindly denies that Parfit's life has been wasted and makes keen observations about Parfit's rejection of internalism, suggesting a more fine-grained survey of the moral field, between 'very strong, robust' and 'very pale, weak' akin to nihilism. The volume contains many other valuable contributions, which Parfit partly incorporates in the third volume of the work, producing a new synthesis, following Temkin's suggestion (2017b, 54). The last volume is in fact largely a companion piece to Singer's collection and not the other way around, as we might think. Schroeder also points out that Parfit's life has not been wasted, despite fundamental doubts about the cogency of his moral theory. Parfit nonetheless re-asserts that 'naturalism, non-cognitivism' and 'subjectivism' had to be fundamentally mistaken: "[...] either we do have reasons to care about some things, or we don't" (2017b, 53).

We must grant that when he inquires in such urgent tones if anything 'really matters', he is deadly serious. Starting with *Reasons and Persons*, he maintains a hope of convincing the 'moral sceptics'. As he asserted then, "[...] one of our deepest disagreements is between Consequentialists and those who believe in Common-Sense Morality" (1984, 114) a disagreement he believed he was able to reduce already, and that if entirely removed by the unified or triple theory, would undermine the case of the Nietzscheans among us. The token to evaluate Parfit's theory should be his own.

Several of the essays respond to Parfit's arguments *against* moral naturalism – in particular his contention that if naturalism were true, moral claims could not state substantive truths. Parfit aims at what he calls 'non-analytical' naturalism, for no moral claim is substantive unless it ascribes an 'irreducibly normative property'. Where do the contributors to Peter Singer's volume think he goes 'wrong' (Simon Blackburn in Singer 2017, 81-98)? Most recent papers address the question of 'naturalism' and of 'substantivity' in different ways and from distinctive viewpoints, namely Peter Railton, Frank Jackson and Mark Schroeder, Allan Gibbard (and Bruce Russell). Since a naturalist does not believe in irreducible normative properties, the specific list of contributors to Singer's

and Kirchin's volume acts in defence of some form of analytical or non-analytical naturalism. They remain unconvinced by the triviality objection. 'Trivial' because according to Parfit, they trivialize the fundamental moral discovery: that some reasons are decisive or compelling and we should follow them.

What remains after Parfit's physical departure? Schroeder comments on *On What Matters* volume I: "[...] if we are evaluating the core arguments of part 6 by the light of this understanding of its overall significance, though it helps us to understand a great deal about Parfit's motivations, it casts suspicion on his methods" (2011). A recent reviewer of volume III, Andrew Sepielli, notes "[t]he rejection of [...] Parfit's arguments by so many authors [...] seems to me to represent the disciplinary consensus" (2018).

We may, however, suggest a different method for drawing a balance. The question is not if the contention that all moral theories have, say, five main components, is correct. The question, given that all moral theories must assume some simplifications, is how many miles you can run with a given set of assumptions. In Parfit's case, many miles indeed. Expressions like 'our duty towards future generations' are current in public discourse for justifying either the Iraq War, retirement pensions reform, sovereign debt or the fight against climate change. Let us pick one example. In 2006, a *Handbook of Intergenerational Justice* gathered together a few numbers to show the increase in interest in justice between generations: in 1980, of over half a million PhD dissertations just one contained 'future generations' or 'posterity' in its title. The situation was already very different ten years ago. The numbers are obviously still behind as regards inequality or poverty, or gender and race, but the gap is closing at a very fast pace (Tremmel 2006, 1). Few dissertations (or books) entirely ignore Parfit's work on these themes, so he set the agenda for much incoming ethical practical/philosophical work, on say pension systems and environmental ethics. Not a small feat, even if the theory really needs (according to non-cognitivists and naturalists) more twists and more new bolts than mere fine tuning. Parfit's claims about future generations became widely influential, since at the least several important moral and political problems, such as the ageing population and some long term environmental issues, can hardly be solved within a self-interested framework, let alone a temporally neutral one.

What needs to be done? Parfit's apparent naïve brand of 'realistic' moral cognitivism is at the root of the enduring scepticism. Complex arguments, refuted assertions, real and improbable counterexamples, surprising theses and stubborn convictions, puzzles, abstract and very real conditions, challenges to the discovery of new theses and surprising conclusions are part of the current analytical philosopher's ecumenical paraphernalia (see Nozick 1974). There is a saying: "If you give me an example, I will give you a rebuttal." Parfit gave many examples, especially because some basic notions cannot be defined, except by making others engage in imaginative intuitions. But are our intuitions not in part shaped by our culture? At the root is Parfit's peculiar method: "[...] in this method common-sense morality is in the driving seat and the main use to which reason is put is in sharpening and refining it" (Kirchin 2017, 20). But it is a derivative common-sense, coming after centuries of modern moral philosophy.

Maybe we cannot merely appeal to moral intuitions while ignoring how things really are, in natural terms. Temkin, among others, points out Parfit's fundamental weakness. He never answers the question of how we can respond, or act moved by non-natural reasons. We think, and most essays in recent literature seem to point, if not by using the same words, that there is an important missing supporting piece in Parfit's triple theory. Without it the success of his arguments would be "a pyrrhic victory for objectivism" (Singer 2017, xi) since moral reasons lack – and cannot have – moving (causal) force over the natural world. That is, we keep stumbling on the missing leg to the triple theory that would make it a more stable device.

This may come from a metaphysical view of what normativity really is about. Perhaps some notion of humanity cannot be entirely avoided and we should stop pretending to know less than we know, just to avoid metaphysical pains. Most opposition to Parfit's theories seems to arise from the suspicion that he tries to re-introduce within analytical philosophy (in veiled form) a hidden metaphysics, as mythical as the beliefs in witches, unicorns and human rights. Parfit denies the need. His ethical theory is one version of non-ontological non-naturalism. He stays away from ontological claims. His method does not make it possible to decide. Should he have stayed away from them for the sake of consensus seeking? He does seem to go further than puzzles, paradoxes and word analysis allow. After all, if meta-ethics is much too hard, how much harder can metaphysics be? Several chapters in *On What Matters* have metaphysics and ontology as chapter titles and Parfit wrote on the most difficult one: "Why anything, Why this?" What needs to be done? Parfit never claimed his work was complete. In his final volume, little is said of 'substantive' normative questions that were the original fulcrum when he was interested in climbing the common mountain of ethics and before thinking that most disagreements were mostly due to lack or misunderstanding of the basic moral categories (see 2017b, part 10).

At 74 years of age, Parfit was certainly one of the most idiosyncratic moral philosophers then alive. If he was right, his personal identity is not what matters, but the Relation (R) between the different states in time (1984, 217; chapter 12). Thus, as Jane O'Grady (2017) wrote in Parfit's obituary and to maintain one of his famous examples, his body may very well have been scanned, a blueprint plotted and teleported to Mars, where it is unclear if his replica is still him, but his personal survival is not what matters (1984, 199ff). But some things *do* matter. Following his path, the question of how we can understand, respond and eventually act driven by non-natural reasons needs to be fully addressed.

## WORKS CITED

- Bratman, M. E.. 2007. *Structures of Agency*. Oxford: Oxford University Press.
- Descartes, René. 1927a. "Meditations." In *Descartes Selections*. Edited by Ralph M. Eaton, New York: Scribners.
- Descartes, René. 1927b. "Reply to Objections." In *Descartes Selections*. Edited by Ralph M. Eaton. New York: Scribners.

- Edmonds, David. 2017. “Derek Parfit obituary.” *The Times*, January 4. <https://www.the-times.co.uk/article/derek-parfit-xvixn0zq5> [accessed January 16, 2018].
- Eaton, R. M. 1927. *Descartes Selections*. New York: Scribners.
- Fenwick, Cody. 2017. “Derek Parfit.” *New York City Patch*, January 2. <https://patch.com/new-york/new-york-city/derek-parfit-renowned-philosopher-ethics-mind-metaphysics-dies-74>. [accessed January 16, 2018].
- Gassendi, Pierre. 1964. *Opera Omnia*. Stuttgart: Friedrich Frommann.
- Gassendi, Pierre. 1927. “Vth. Objections?” In *Descartes Selections*. Edited by Ralph M. Eaton, 245-246. New York: Scribners.
- Gibbard, Allan. 2017. “Parfit on Normative Concepts and Disagreement.” In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 61-80. Oxford: Oxford University Press.
- Gilson, Etienne. 1937. *The Unity of Philosophical Experience*. New York: Scribner’s.
- Huddlestone, Andre. 2017. “Nietzsche and the Hope of Normative Convergence.” In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 169-194. Oxford: Oxford University Press.
- Husserl, Edmund. 1959. *Erste Philosophie* (1923/4), vol. 8. The Hague: Martinus Nijhoff.
- Jackson, Frank. 2017. “In Defense of Reductionism in Ethics.” In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 195-212. Oxford: Oxford University Press.
- Kirchin, Simon (ed.). 2017. *Reading Parfit. On What Matters*. London: Routledge.
- de Lazari-Radek, Katarzyna and Peter Singer,. 2017. “Parfit on Objectivity and ‘The Profoundest Problem of Ethics’.” In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 279-296. Oxford: Oxford University Press.
- McMahan, Jeff. 2017. “Derek Parfit (1942-2017).” *Philosophy Now* 119. [https://philosophynow.org/issues/119/Derek\\_Parfit\\_1942-2017](https://philosophynow.org/issues/119/Derek_Parfit_1942-2017) [accessed January 16, 2018].
- Mayer, Christian. 1993. *The Political Art of Greek Tragedy*. Baltimore: Johns Hopkins University Press.
- Molière. 1871. *Oeuvres Complètes de Molière*, vol. 2. Paris: Garnier.
- Nagel, Thomas. 1986. “Personal Identity and Reference.” In *The View from Nowhere*. Edited by Thomas Nagel, 37-43. Oxford: Oxford University Press.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Oxford: Blackwell.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge: Harvard University Press.
- Offray de La Mettrie, Julian. 1996. *Machine Man and Other Writings*. Cambridge: Cambridge University Press.
- O’Grady, Jane, 2017. “Derek Parfit Obituary. Philosopher Whose Books Inspired his Academic Peers All Over the World.” *The Guardian*, January 12. <https://www.theguardian.com/world/2017/jan/12/derek-parfit-obituary> [accessed January 16, 2018].
- Olson, Eric T. 1999. *The Human Animal*. New York: Oxford University Press.
- Olson, Eric T. 2003. “Was Jekyll Hyde?” In *Philosophy and Phenomenological Research* 66: 328-348.



- Palmer, David. 2014. *Libertarian Free Will*. Oxford: Oxford University Press.
- Parfit, Derek. 1971. "Personal Identity." *Philosophical Review* 80: 3-27.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press [reprinted with minor corrections in 1985, 1987].
- Parfit, Derek. 2008. "Persons Bodies, and Human Beings." In *Contemporary Debates in Metaphysics*. Edited by Theodore Sider *et al.*, 177-208. Oxford: Blackwell.
- Parfit, Derek. 2011. *On What Matters*. Volumes I-II. Oxford: Oxford University Press.
- Parfit, Derek. 2017a. "Responses." In *Reading Parfit. On What Matters*. Edited by Simon Kirchin, 189-236. London: Routledge.
- Parfit, Derek. 2017b. *On What Matters*. Volume III. Oxford: Oxford University Press.
- Plantinga, Richard J., Thomas R. Thompson, and Matthew D. Lundberg. 2010. *An Introduction to Christian Theology*. Cambridge: Cambridge University Press.
- Quine, Willard van Orman. 1972. Review of Milton K. Munitz (ed.) *Identity and Individuation*. In *The Journal of Philosophy* 69: 488-497.
- Railton, Peter. 2017. "Two Sides of the Meta-Ethical Mountain?" In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 35-60. Oxford: Oxford University Press.
- Bruce Russell. 2017. "A Defence of Moral Intuitionism." In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 231-258. Oxford: Oxford University Press.
- Rudder Baker, Lynne. 2000. *Persons and Bodies*. Cambridge: Cambridge University Press.
- Ryle, Gilbert. 2009. *The Concept of Mind*. London: Hutchinson's University Library.
- Scheffler, Samuel. 2013. *Death and the Afterlife*. Oxford: Oxford University Press.
- Schroeder, Mark. 2011. "Derek Parfit. On What Matters, Volumes 1 and 2." <http://ndpr.nd.edu/news/on-what-matters-volumes-1-and-2/> [accessed January 18, 2018].
- Schroeder, Mark. 2017. "What Matters About Metaethics?" In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 213,230. Oxford: Oxford University Press.
- Sepielli, Andrew. "Derek Parfit, On What Matters, Vol III." <http://ndpr.nd.edu/news/does-anything-really-matter-essays-on-parfit-on-objectivity/> [accessed January 18, 2018].
- Shoemaker, Sydney. 1968. "Self-Reference and Self-Awareness." *Journal of Philosophy* 65: 555-567.
- Sidgwick, Henry. 1907. *The Methods of Ethics*. London: Hackett Publishing [reprint 1981].
- Singer, Peter (ed.). 1985. *In Defense of Animals*. Oxford: Basil Blackwell.
- Singer, Peter (ed.). 2017. *Does Anything Really Matter? Essays on Parfit on Objectivity*. Oxford: Oxford University Press.
- Temkin, Larry. 1996. *Inequality*. Oxford: Oxford University Press.
- Temkin, Larry. 2017. "Has Parfit's Life Been Wasted?" In *Does Anything Really Matter? Essays on Parfit on Objectivity*. Edited by Peter Singer, 1-34. Oxford: Oxford University Press.

- Tremmel, Jeorg Chet. 2006. *Handbook of Intergenerational Justice*. Cheltenham: Edward Elgar.
- Williams, Bernard. 1970. "The Self and the Future." *Philosophical Review* 79/2: 161-180.
- Williams, Bernard. 1973. *Problems of the Self*. Cambridge: Cambridge University Press.
- Wittgenstein, Ludwig. 1958. *The Blue and Brown Books*. New York: Harper and Row.