Parfit and the Buddha: Identity and Identification in *Reasons and Persons*   by Robert Ellis, Lancaster University

[published in *Contemporary Buddhism,* May 2000]

This is an attempt to examine some of the most important arguments of a leading contemporary moral philosopher, not so much *in comparison* with Buddhism, as *from a Buddhist point of view*. Derek Parfit's Reasons and Persons (1984) goes a long way, from a fairly traditional analytic starting point, towards some useful conclusions which attempt to engage with issues of great contemporary importance in ethics and the philosophy of identity. In this paper, though, I shall argue that Parfit's arguments about identity, rationality and impersonality in ethics are much less clear and useful than they might be. This is because of an unacknowledged tension between what might broadly be called the Buddhist and the analytic tendencies in his work. Whilst Parfit claims that the Buddha would have supported his views on personal identity, I shall claim that the resemblance is superficial because Parfit's apparently Buddhist conclusions depend on much narrower analytical premises. The discrepancy centres particularly around the difference between Parfit's concern with identity as opposed to the Buddha's concern for identification, and I shall argue that the Buddha's approach leads more clearly into the kind of ethical implications that Parfit wants to draw from his arguments about identity.

1.   Self-identity and self-identification
a.   Parfit and the Buddha

I shall begin with what is probably the most fundamental concern of *Reasons and Persons*: that of the nature of persons. Parfit examines psychological, physical and Cartesian criteria for personal identity but concludes that the question of what correct criteria should be is an empty one. There is no determinate answer to the question 'Will it be me?', only a conventional one. Personal identity, he claims, contrary to our instincts, is not "what matters": instead, he claims that what matters is psychological continuity and connectedness, even when caused by unusual means such as teletransportation. In fact we have no grounds for rational worry, he claims, when we step into a machine that will annihilate our body and recreate a qualitatively identical one somewhere else.

Parfit's reductionism here bears a superficial resemblance to the Buddhist doctrine of no-self or anatta, but for the reasons I shall explain Parfit misrepresents the Buddha in claiming that he would have agreed with Parfit's views. The underlying difficulty is that whilst both Parfit

and the Buddha reject fixed criteria for self-identity, Parfit does so on grounds of metaphysical truth whilst the Buddha does so on pragmatic grounds[1]. This distinction is perhaps clouded by the fact that Parfit *does* want to apply his conclusions about personal identity in the ethical sphere, in order to support the impersonality of utilitarianism. However, despite this instrumental intention his argument about personal identity is based on grounds of independent fact. The Buddha is not concerned to enter into an ultimately fruitless debate about whether or not there is an absolute criterion for self- *identity*, but to discourage self-*identification*. To do this it is not necessary to make or deny any metaphysical claims, but only to show the harmful effects of self-identification, even according to the values implied in this identification.

Self-identity, or its absence, is a matter of fact. Self-identification, on the other hand, is a matter of desire and belief. Whether or not it *is* the case that the 'me' who gets up tomorrow will in fact be me, today, typing this, it is certainly the case that I *want* to wake up tomorrow. I imagine *myself* waking up tomorrow, and want it to be me. I also want a number of other things to continue in an identical fashion which I identify with myself: I want my body to be intact and my possessions to be still functional and not stolen, and I want my mind and reputation to be unimpaired. All of these continuities are thus functions of the ego. According to Buddhist teachings, however, this identification is not inevitable and can be lessened and ultimately removed, by spiritual practices. The purpose of a disbelief in determinate self-identity then, is a wholly instrumental one and is not dependent on any idea of the truth or falsity of that belief independent of its instrumentality.

I shall argue that Parfit's position is inconsistent and incomplete in comparison to the Buddhist one. Parfit does not have sufficient grounds to assert reductionism as a factual metaphysical claim, yet he seems to assume that it is necessary to do so in order to fulfil an instrumental purpose similar to the Buddha's. He appears to deliberately avoid any discussion of the causes of our belief or disbelief in the self, and yet he also believes in the moral usefulness of not believing in it, saying that he finds liberation from belief in the self personally both 'liberating and consoling' (p.281). Parfit mentions, but does not explore any explanation for, the contrast between this reaction and Hume's (who found it depressing), providing little more than the hint that 'the unity of our lives is a matter of degree, and is something we can affect' (p.447). As I shall argue, an account of how this unity is to be achieved is more important as a basis of ethics than any ontological assertions about the self.

b. Why we have no reason to accept the analytic account of the self

Parfit might argue here that his restraint is one merely in harmony with the proper bounds of discussion within the analytic tradition of philosophy, and although he might feel some sympathy with a Buddhist account of non-egocentric moral development, to give an account along those lines would involve him in making unjustifiable psychological assumptions. Among these assumptions would be that of the existence of a self-creating, self-identifying ego. This might be seen merely as a construction of empirical psychological theory.

In any discussion of the self, however, some psychological presuppositions are inevitable. The analytic tradition of discussion about personal identity tends to take *cognitive* awareness as its point of departure, debating whether or not momentary self-awareness necessarily implies the existence of a distinct mental entity. The focus is thus on conceptualising our experience in terms of objects of knowledge. Any momentary self-awareness, however, will also include affects: some kind of reaction to stimulus must be present in the form of some interest or desire which forms the basis of continued consciousness, even for the thinker in a classic Cartesian state of abstract reflection without any accompanying action. 'Ego' is the term I shall use for the implied subject of this affect, just as 'self' or 'soul' may be used as the implied subject of cognitive consciousness (I mean 'ego' in a sense descended from Freud: Parfit confusingly talks about the 'Cartesian ego'). Unlike the self, the ego does not have to 'exist' over time: it consists only in a collection of desires existing at any given instant.

Both of these terms are theoretical constructions which can be used as the basis for further reasoning. I cannot see why there is any more or less reason *a priori* to accept the cognitive self (or its absence) than there is to accept the affective ego (or its absence). The cognitive seems no more or less basic to my experience than the appetitive. I would suggest that the only way to draw any clear conclusions on the truth of these metaphysical constructions is through their heuristic value. If either construction, combined with auxiliary hypotheses, proves more fruitful than the other we begin to have more reason to accept it. Part of my position is the claim (which I do not have space to fully defend here) that to provisionally assume the existence of an affective ego is much more *ethically* fruitful than either to assume its non-existence or to assume the existence of the cognitive self. Since the cognitive self is a fixed quantity whilst the affective ego is dynamic[2], an assumption of the ego's existence also needs to carry with it an assumption either of the non-existence of the self or of its complete irrelevance to ethics, so as to avoid a contradiction between self and ego when they are

applied to ethical theory. In denying determinate criteria for personal identity, Parfit denies the self, but he does not at the same time affirm the ego.

c.   The impossibility of impersonal description

Parfit's basic description of the reductionist position (p.210), which he later goes on to advocate, rests on the possibility of an impersonal description of personal identity. As he writes:

> …the fact of a person's identity over time just consists in the holding of more particular facts…. These facts can be described without…presupposing the identity of this person…. These facts can be described in an *impersonal* way.

The possibility of such an impersonal description appears to depend on the view one takes of the relationship between objective facts and the description of facts. If there is an absolute one-to-one correlation between symbols and an atomised reality, in the way envisaged by the Wittgenstein of the *Tractatus*, it appears that description could be wholly impersonal: but such an ideal does not appear to bear any relationship to any conceivable use of any conceivable language. One could imagine a brain surgeon probing Descartes' brain and describing his brain-processes in an apparently wholly impersonal manner as he thinks 'Cogito ergo sum.', but such a commentary would still be composed of subjects and predicates which presupposed the identities of smaller units. The language employed might treat Descartes and his identity solely as an object, but still be subject to the personality of the brain surgeon and his assumption that Descartes and his brain-processes were <u>not</u> part of him.

The impossibility of an impersonal description can thus be shown in two ways: (i) from the fact that an ideal reductionist language would merely reduce a larger substantive identity into smaller ones, which would then have to be further reduced themselves, leading to an infinite regression of reduction. Any part of Descartes may conceivably contain the soul of Descartes, and its operation must be described in the same subject-predicate terms as Descartes' actions as a whole person. (ii) from the fact that the person employing language must assume her own identity in order to make an 'impersonal' description. In order to describe a person's operations in objective terms she must distinguish those operations from her own. The description can therefore not take place in a universe which is theoretically free of subjects, since, even if there was only one subject left, she could not deconstruct her own operations into reductive language.

From the impossibility of verbal reductionism, however, we are not entitled to deduce the impossibility of ontological reduction. It may be that in fact I am nothing more than the sum of my parts, but the fact that I am also an ego makes it impossible for me to describe myself in those terms. Nor does the impossibility of verbal reductionism imply anything about the likelihood of ontological essentiality. What I believe this impossibility to show is the continual presence of self-identification wherever the conceptualising properties of language operate, another indication of the operation of the ego.

d.   Parfit's S and the ego

Before he reaches his reductionist account of the self in Part 3 of *Reasons and Persons*, Parfit has already given a parallel argument about rationality which deals not with the ontological status of the self but the rational coherence of a theory which takes self-fulfilment as its aim. Parfit's S, however, is not the egoistic impulse as we experience it: rather it is a theoretical position which attempts to systematise egoistic motivations. Parfit argues that whether a hedonistic, desire-fulfilment or 'objective list' view of the self's proper goals is taken, a follower of S ought to seek his proper goals in a temporally neutral fashion. S is thus in opposition to three egoistic biases which the S-theorist regards as irrational: the bias to the near (as opposed to the further future), the bias to the future (as opposed to the past) and the bias to the present (as opposed to the past or future).

This conflict between S and the ego immediately raises difficulties. The S-theorist appears to be claiming that what is good for me is neither what I identify with and desire, nor a morality which I do not identify with. What is it, then, which gives the goals of S value? Parfit seems to assume that it is logically possible for an outcome to be better for me according to criteria which are neither my own nor beyond me. This outcome has an anomalous status which mirrors that of my replicant on Mars. If I do not acknowledge this outcome as better, in what sense is it better *for me*? For an outcome to matter *for me* it seems that it must be connected to my concerns, just as to *be me* an entity must be connected to what I am concerned about.

This raises the question of the distinction between rationality and morality, which I shall discuss in the next section. For the moment, however, it must be asked whether Parfit is attacking a straw man when he attacks S. If S does not represent a position which anybody is concerned about, then defeating it does not show why anyone should be any less egoistic.

Nevertheless I believe some aspects of Parfit's strategy in attacking S to be useful. If instead of taking S to be a wholly abstract theory we took it to be the range of self-identification of any given ego, the strategy that Parfit uses in attacking S could be just as effective in attacking egoism. Briefly put, Parfit's strategy is to contrast S with both Present-aim theory (P) and with morality, and then show that S's defences in each case are inconsistent. Against P, the S-theorist must argue for temporal neutrality, but against morality the S-theorist must argue for personal bias. Parfit seems to have a strong case for arguing that S, as given, is inconsistent in advocating bias in one case and neutrality in the other, especially if this is supported by the later argument that less unity of the self implies less disunity between selves.

However, the picture is clarified further if we substitute the ego for S. The ego consists of a variety of desires which are not necessarily rationally coherent, since the desire for continuity may be attached to many contradictory objects (e.g. both objects and ideologies which criticise the possession of those objects). At any one given time our present aims only encompass one desire or coherent set of desires. To defend any given ego, then, we would have to argue that not only the desires of any given moment (however rational these may appear) should be given priority for fulfilment, but the desires which manifest themselves over a period of time. This could, perhaps, only be done through some sort of appeal for distributive justice among desires. If distributive justice between the desires of one person is to be advocated, though, why not between the desires of all persons?

In this way the incoherence and irrationality of the ego can be revealed even on the basis of its own assumptions (though the rationality appealed to is non-egoistic). Parfit has an argument which is formally excellent but misapplied due to the over-narrow analytic bounds within which he has conceived his task.

e.   S and the ego in relation to time
A further problem arises when an egoistic version of S is compared to Parfit's version in relation to time. This is because the time that we experience is projected from the present and thus must be conceptualised by the ego. When I experience a memory or an anticipation and am aware of it as such I must simultaneously label it 'past' or 'future', despite the fact that the memory or anticipation occurs in the present. Whatever events I recall or anticipate are in some way already identified with merely through the fact that they are recalled or anticipated.

This follows from the argument I have already given in §1c above regarding the impossibility of impersonal descriptions. Our conceptions of past and future must consist at least partially of descriptions or they are not acknowledged as such, being merely experiences in the present. Such descriptions are necessarily personal in nature (in the sense that they adhere to the subject-object dualism of language) and therefore presuppose an egoistic view to some degree.

This implies that Parfit's P and S (as well as M, C, N, and all the other theories he mentions) are all in fact indistinguishable with regard to time. The 'temporal neutrality' of S is a way of describing a present intention to take into account past and future desires and apply distributive justice between them. Parfit in fact rightly concludes (p.135-6) that the issue between S and P is not one of temporal neutrality at all, but that whilst S is necessarily egoistic, P is not (because present aims could also be altruistic or neutral). Our present aim could, he in effect points out, be identical in content to S (or not). However, Parfit draws this conclusion for reasons which are inconsistent with his earlier depiction of S in purely theoretical , not in egoistic terms. Parfit's S is not necessarily egoistic in the sense that it follows the desires of any given ego: but to say it is egoistic in the sense of following S is trivial. P and S thus appear totally indistinguishable both in terms of time and in terms of necessary egoism, and no useful point has been made.

A much more useful way of, once more, applying the form of Parfit's argument but changing the content is to substitute the ego for S, where the ego consists in a number of desires which exist *at the same time*. So that the ego is not identical to P we must also substitute for P a single desire amongst the strand existing at one time, which I shall instead call D. Desires are not necessarily simultaneous in the sense of occurring in the same consciousness at the same time, so we must assume for this purpose the existence of unconscious desires which exist in a potential form simultaneous with conscious ones. This is an assumption which may raise other philosophical issues, but, like the assumption of the ego, I think this assumption may be justified on heuristic grounds by its fruitfulness in avoiding the circularity of Parfit's account. On this new account what Parfit describes as 'temporal neutrality' can be better explained as a distributive justice between all the desires D existing at one time. There is now no doubt about the distinctness of D and the ego, but the conflict between them can by exploited to undermine egoism in exactly the form I outlined in §1d above, utilising the inconsistency between the egoistic response to D and the egoistic response to morality.

It may be asked whether the ego in fact operates any such distributive justice, or whether I am simply constructing another abstract position liable to the same difficulties as Parfit's S. The difficulty is a real one, though not exactly the same as that of 'temporal neutrality'. At any one given time the ego merely identifies with the desire or set of desires D of which it is conscious, giving D priority over other desires and so in this sense not operating any distributive justice. From the viewpoint of the present the ego appears to be identical to D. However, it can also be argued conditionally that if the ego were equally conscious of all desires D at that time, it would identify with all of them and take them into account in proportion to their strength, thus operating a sort of distributive justice. We have to assume this conditional point as an extrapolation of the fact that when consciousness of hitherto unconscious desires does occur, those desires are immediately identified with. By making this conditional construction, however, we can avoid the difficulties created by the lack of a determinate criterion of identity, due to which desires emerging in the past or the future may be quite different (though probably continuous with) those which exist unconsciously in the present.

So far I have tried to show a number of ways in which Parfit is attacking the wrong target (even when 'wrong' is seen in terms of his own apparent intentions and values), when he attacks self-identity rather than self-identification. The central reason for this is that self-identity is completely irrelevant to ethical concerns unless it coincides with self-identification. This, however, has already raised the issue of rationality. Does it make sense to discuss an egoistic rationality of the type which S represents? This is the subject of the next section.

2. **Rationality and Morality**

a. **Parfit's contradiction**

Parfit devotes a good deal of space in Parts 1 and 2 of *Reasons and Persons* to explaining the basis of an egoistic rationality: that is, that it may be rational to think and act in a way which is in one's own best interests and yet neither necessarily in accordance with ones immediate desires nor morally. His advocacy of a critical version of P (CP) as a more rational substitute for S depends on this conception of rationality. His explanation of CP goes as follows.

> CP: Some desires are intrinsically irrational. And a set of desires may be irrational
> even if the desires in this set are not irrational….Suppose that I know the facts and am
> thinking clearly. If my set of desires is not irrational, what I have most reason to do is
> what would best fulfil those of my present desires that are not irrational. This claim
> applies to anyone at any time. (p.119)

CP can 'appeal to the agent's values, or ideals, or to his moral beliefs' (p.119) but does not necessarily do so, and thus is quite distinct from morality.

To support this Parfit attempts to give criteria for desires which are intrinsically irrational, giving a number of examples (p.120-126). For example, he claims that a person who was altruistic towards people who lived less than exactly a mile away, but not towards people living more than a mile away, would be irrational (p.125). He then summarises the criteria.

> In these cases the concern is not less because of some intrinsic difference in the object of concern. The concern is less because of some property which is purely positional, and which draws an arbitrary line. These are the patterns of concern which are, in the clearest way, irrational. (p.126)

The weaknesses in this account show the precariousness of this whole concept of non-moral rationality. Why is the 'within-a-mile altruist' so irrational? This person, concluding perhaps that it was not feasible to be altruistic towards everybody, has adopted a consistent programme of whom to be altruistic to. By the same criteria Parfit would have to condemn as irrational the enforcement of most laws, which depend on the apparently arbitrary drawing of lines of concern: speed limits are an obvious example. The distinction between 'intrinsic' and 'purely positional' is of course an entirely conventional one which itself draws an apparently arbitrary line. If it is not clear which desires should be seen as irrational, then there is no basis for CP, and no theory which can take the place of S as giving an account of non-moral rationality.

In making his points about rationality Parfit appeals to an essentialism which is clearly contradicted in his discussion of identity in Part 3, where he claims that determinate criteria of identity are not 'what matters' and thus essentialism is denied. He further goes on to explicitly deny that we should see rationality and morality as distinct, in order to explain why we should condemn imprudence when there are no determinate criteria of self-identity (p.318-320). As far as I can see, Parfit directly contradicts himself here. The implication of there being no determinate criterion for self-identity appears to be that CP, which depends on criteria for rationality which appeal to intrinsic features, is just as rational or irrational as S, which depends on inconsistent criteria for identity over time. Parfit's reductionism cannot be sustained without an implied nominalism which undermines the recourse to essentialism in his theory of rationality.

b. An alternative view of rationality

As previously, I believe that Parfit's difficulties are due to his over-concern with self-identity rather than with self-identification. Just as it is difficult to sustain a real difference between S and P, so it is difficult to maintain a distinction between egoistic rationality and desire. The attempt to describe what is in my interests distinct from what I want appeals neither to the desires of the ego nor to any objective or impersonal criterion, so it is very difficult to see why we should value what it says we should value except for moral or egoistic reasons.

As an alternative view I want to argue that there are only two types of rationality, though the distinction between them is a matter of degree. There is the rationality of the ego, which consists in nothing other than what I desire, both consciously and unconsciously, at the present moment; then there is objective rationality beyond the ego. The rationality of the ego is bounded by its subject-object dualism, its dependence on conceptualisation, and its use of determinate criteria of identity which are in fact conventional. For this reason it might be referred to as *conventional rationality*. However, the degree of awareness of the conventionality of these determinate criteria accompanying a given desire is a matter of degree, so that progress can be made towards the *absolute rationality* beyond the ego without strictly speaking going beyond the bounds of conventional rationality.

As I have already argued, the desires of the ego may be inconsistent. Even where they are not internally inconsistent, they are necessarily inconsistent with the desires of other egos (if they were not, one ego would not be distinguishable from another and would thus no longer exist)[3]. Analysis like the modified form of Parfit's argument that I have given above shows that the 'rationality' of the ego is no rationality at all, since from a transpersonal or objective point of view it is self-contradictory. Only a single desire (or set of coherent desires) D, taken in isolation, could possibly be seen as not self-contradictory: but we do not in fact experience any such isolated desires.

So in this sense we are all as 'irrational' as Parfit's within-one-mile altruist, creating arbitrary boundaries within which we try to assemble some consistency. The setting of any such boundaries creates a sorites paradox whereby other possible boundaries adjacent to the one chosen may seem just as rational. Insofar as there is consistency within these conventional boundaries, however, there is a degree of rationality. Degrees of rationality are thus in inverse ratio to degrees of self-identification. Rationality in this incremental sense is also identical to morality.

This alternative scheme allows me to make more sense of Parfit's claims about CP. Parfit has written that the critical criteria applied to judging present aims in CP could be moral ones (though they need not be necessarily), and that for this reason he thinks moral theorists should accept CP (p.194). If 'morality' is understood in the sense above and the option is turned into a requirement, CP could be understood as a broad description of the incremental application of absolute rationality to our present desires. The important point Parfit has clearly made, then, is that rationality cannot lie outside the sphere of what we actually desire.

c. 'What matters'

Parfit's apparently inconsistent view of rationality also occurs in Part 3. Here, after arguing that a determinate criterion of personal identity does not matter, he also claims more dubiously that 'what matters' is psychological continuity and connectedness howsoever caused, making it irrational to fear teletransportation. The reasoning Parfit gives for this claim is that it is irrational to care about the cause of psychological continuity – one should only care about the effect (p.286).

My worry in such a case would not be about the cause of my teletransportation (assuming the machine to be reliable), but about the effect. I would be worried lest the effect of the process would not be me: the reason for this being, as I have already indicated, that I would not identify myself with my qualitatively identical replicant. The only circumstances in which I might conceivably identify myself with the replicant might be if I had been indoctrinated to believe that the replicant was me by being brought up in a society where what Parfit calls 'series persons' are the norm of human life. What my ego identifies with is completely contingent, but the fact remains that, where it does not identify with my replicant, I shall be seriously worried.

On what grounds of rationality might I be worried? Certainly not on the absolute grounds on which Parfit argues that determinate self-identity does not matter. If I was absolutely rational I would have no cause for alarm. However, within the restricted rationality of the ego, which models its own, contingent, view of what is or is not worthy of my concern, I would have very good grounds for worry.

The difficulty with Parfit's two claims is thus that he makes them on inconsistent grounds. The claim that absolute self-identity does not matter is made on absolute grounds whilst the

claim about what does matter is made on conventional grounds, presumably (though this is not made very explicit) of CP rationality. Unless I share those conventional assumptions I have no reason to accept the latter claim. A further difficulty is that Parfit does not appreciate that the 'rationality' of the ego arises out of a physically and emotionally grounded consciousness and is not purely cognitive. It is for this reason that egoistic and absolute claims are often incompatible.

3. <u>Impersonality</u>

a. **Identification and disidentification**

In the final section I want to look at what Parfit gives as the most important common feature of the arguments in his book: impersonality. As he writes in his concluding chapter:

> I have argued that, in various ways, our reasons for acting should become *more impersonal*. Greater impersonality may seem threatening. But it would often be better for everyone. (p.443)

He goes on to argue that some of the 'five mistakes in moral mathematics' he has offered (e.g. ignoring the effects of sets of acts, ignoring small chances, and ignoring imperceptible effects) did not matter in the small communities of traditional societies, where the effects of actions were more obvious and limited in extent. He then goes on to argue:

> Life in big cities is disturbingly impersonal. We cannot solve this problem unless we attack it in its own terms. Just as we need thieves to catch thieves, we need impersonal principles to avoid the bad effects of impersonality. (p.444).

These two quotations for me reveal both the admirable and the mistaken aspects of Parfit's argument. I will begin with the admirable. The assumption of a consequentialist perspective allows him to diagnose clearly some of the mistakes in moral argument which have created huge problems in the modern world, particularly those involving the use of resources and the attitude to the environment. He has done this by adopting a highly impersonal perspective, and for him his arguments about S and his advocacy of reductionism are contributions to the argument as to why we should adopt such a perspective. The reductionist view, particularly, he argues, should lead us to take a utilitarian rather than retributivist view of crime, focus more on experiences than on persons as the basis of moral value, and take imprudence to be morally wrong.

In my view these are generally helpful conclusions, despite the fact that they have been reached through a dubious argument. But, as I have argued, Parfit has drawn these conclusions on the basis of an ontological claim about personal identity which does not

justify any moral implications unless we are also given *reasons* for disidentification with the ego. We will do the things he recommends *if* we succeed in disidentifying from the ego, but without such a disidentification we have no reason not to continue to think in personal terms.

Parfit's recommendations thus come from an absolute perspective which allows us to clarify the issues but does not, by itself, produce a moral solution to them unless the psychological conditions are also taken into account. Parfit's argument that we need impersonality to deal with impersonality seems to be psychologically naïve. Given that, as I have argued, identification with the self is created by the ego, moves towards impersonality argued from a perspective which does not take into account the nature of our resistance to it are only likely to create resistance and conflict. From the designation of impersonality as the ideal state does not follow its usefulness as a moral prescription.

Even if Parfit were to shift his recommendation from one of simply adopting an impersonal perspective to one of disidentification from the ego, this would still not be sufficient to cover the whole process of human moral development. Disidentification cannot take place unless identification has already occurred, and the two are in many ways not opposed or even wholly distinct processes. As Parfit argues, the disunity of the person implies greater unity with other persons, a central point which implies that to *realise* this disunity is simultaneously to realise ones unity with others. This realisation may only be possible on the basis of a degree of prior identification, a conventional basis on which to profitably understand and apply the recognition of our ultimate contingency. Parfit's recommendation of impersonality alone may be positively harmful for those who have not yet achieved such a conventional basis.

The move from pure impersonality to a more complex psychological account of ethics involving identification and disidentification moves us out of the domain of pure consequentialism and into a type of virtue ethics. In this approach to virtue ethics both a conventional basis of identification and ethical refinement through disidentification are dispositions of an individual mind, and ethical objectivity is a product of this virtue rather than being seen in terms of impersonally desirable outcomes in a supposedly objective experience. Despite his adherence to consequentialism Parfit does provide more interesting resources to aid an understanding of a virtue ethics of this type, but again he needs to be re-interpreted in a Buddhist vein rather than in too narrowly rationalistic a fashion.

b. The improvement of theories

One particularly interesting passage of *Reasons and Persons* which provides such resources is his recommendation for the improvement of common-sense morality (M) in order to move it more in the direction of consequentialism (C). Parfit here seems to be interested in the issues raised by a developmental account of the holding of ethical theories, namely why people should come to adopt more impersonal theories. Here also he does not present personal and impersonal in dichotomous confrontation but considers them incrementally. Whilst C operates at an absolute and impersonal level, M may be said to represent a perspective from somewhere within the process of later ego-identification overlapping with early disidentification. M represents a theory in which we have a moral responsibility not just for ourselves but those with whom we have a particular relationship, and we give priority to the welfare of these people over that of strangers (p.95). Parfit argues that M is directly collectively self-defeating, because to act, even successfully, according to this theory is not necessarily in the best interests of the favoured people (p.95-8): this involves merely extending the arguments he has used about self-interest in prisoner's dilemmas to encompass group-interest. For example, fishermen who continue to fish an over-fished water for the benefit of their families will in the long-term work against the interests of their families if all the fishermen do likewise.

As in Parfit's treatment of P and S, points which would be stronger if they were related to the ego are instead made only on the basis of an abstract theory the relative egoism of which is ambiguous. On the one hand Parfit here successfully shows a way in which an egoistic motivation applied to favoured others can produce exactly the same pattern of self-defeat as that of pure egoism, with the further point that when this egoism tries to dissemble itself in the non-egoistic covering of a moral theory, its failure becomes even more obvious. On the other hand, a difficulty arises from the fact that M is not necessarily held egoistically, and may be a theory held by those who are sincerely trying to be non-egoistic: in this case, a deontological theory results. As Robert Merrihew Adams (Adams 1997) points out, the sincere holding of a deontological moral theory may lead to priority being put on avoiding wrongdoing regardless of the consequences: the fisherman may feel it is his duty to feed his children in the present regardless of the fact that they may starve later: for those holding such a theory failure appears to be impossible on their own terms (just as experiential evidence supporting the theory is impossible).

Even if Parfit has not succeeded in showing that M fails in its own terms due to self-defeat, it is clear that it fails in the terms of C. For those who have adopted M for egoistic reasons, C may be too demanding to produce anything other than rejection, so for these Parfit's R – a revised version of M – may be recommended. This revised version of M moves towards C but takes its moral justification from M and does not pretend to lead to optimal consequences. The revision R depends on the contractual idea that everyone should agree not to favour their 'M-related people' provided that a critical mass of at least k people also decide not to do so (p.100-103).

Although on a criterion of judgement which only provides for individual self-interest, it would still apparently not be in the interests of any individual to join this contract, we have to bear in mind that even egoistic motivations involve identification beyond the bounds of the individual, and that any individual holds theory M must indicate at least that they have thus extended their identification, however narrowly and possessively they may be doing so. Parfit argues only that anyone who subscribes to M ought to subscribe to R. He recognises that the premises of R are not exactly those of M (the range of M-related people is being extended to include all those who subscribe to the contract), but argues that 'each *ought* to contribute since, though each is doing what is worse for his own children, the k contributors are doing what is better for all their children' (p.101).

It is not at all clear, however, why all the contributors should suddenly become concerned for other people's children if their range of identification has not been stretched at all. Parfit's argument does not succeed on purely analytic grounds. What it might provide, however, is a good example of the theoretical dimension of the process by which the development of the ego through identification and disidentification might take place. The fisherman who ceases to fish over-fished waters because he comes to identify more with all the children of his colleagues simultaneously ceases to identify so exclusively with his own to the same extent, and he would perhaps never have been able to reach this point without first identifying with his own children's welfare. The heuristic process by which an individual comes to accept new ethical theories involves both the development of new theories and their relative confirmation or falsification through experience. A new theory, if it is likely to be both accepted and enacted, may extend the previously-accepted premises only little by little. In defiance of the sorites paradox, it thus appears to be possible to move incrementally from a narrowly egoistic position to an increasingly non-egoistic one.

c. Impersonality and enlightenment

Finally I want to return more explicitly to the initial theme of comparison between Parfit's approach and the Buddha's by suggesting some ways in which Parfit's ideas, inappropriately impersonal as they are to most ethical considerations, can help to clarify our understanding of what an absolute state of virtue, or enlightenment, might be like. My purpose in doing this is not to try to prove that enlightenment is necessarily possible, but to give it greater consistency as a regulative ideal, so that even if it is considered that it cannot be achieved in its pure state it will be clearer what it means to approach enlightenment. One of the chief difficulties which exist in understanding enlightenment is that of reconciling complete non-egoism with existence as a person. Why does an enlightened person want to continue existing? How do they operate without an ego? I think that Parfit helps a good deal with this through his claim that we should see ourselves, not as persons, but as series-persons.

In the case of the teletransportation example, we may wonder what kind of state is required not to worry about ones fate in teletransportation: the answer must be a completely non-egoistic one where there is no identification with ones body. A person without ego may be *impersonally* concerned about what may be achieved with her body, and for that reason may want the teletransportation to take place without error so that her purposes may continue to be fulfilled on the new planet, but apart from this has no more identification with the body being annihilated than with anyone else's. In this sense Parfit's broad causal relation R may be said to be 'what matters' for an enlightened person. The person being teletransported might be compared to the bodhisattva in Mahayana Buddhism, who is said to deliberately take rebirths in order to continue to help sentient beings until all are brought to enlightenment. Such a person might be described as having impersonal motives personally expressed.

In Parfit's description of series-persons which exist alongside ordinary persons one could also read the Mahayana Buddhist teaching of *Buddha-nature* (see Williams 1989 ch.5), the idea that an enlightened person already exists within us. Take the following passage:

> No phoenix has ever existed. But there are many series-persons. These sentences are being typed by a series-person, me. They are also being typed by a person: old-me. This person is named Derek Parfit. I the series-person hereby name myself *Phoenix Parfit*. Since my present body is also Derek Parfit's body, we are both of us typing these sentences….But….The difference between us is this. On Nagel's view, if Old-I was teletransported, this would kill old-me, the person. But it would not kill me, the

> series-person. This difference is enough to make old-me and me different individuals.
> (p.291)

Parfit of course does not recognise any psychological difference between Derek Parfit and Phoenix Parfit, but perhaps this does not matter. We can simply say that whichever of the two Parfit truly identified with at the time of teletransportation, would be him. Enlightenment is in this way shown to be merely a matter of disidentification rather than a change into something we are not already.

Parfit's idea of quasi-memories and quasi-intentions in cases of split and merged identity also perhaps help to delineate the mental events of an enlightened person. To operate as a person an enlightened person needs all the usual mental events and processes, yet the absence of an ego means that such a person must have them non-egoistically. The ego, however, is what unifies our mental operations such as memory, intention, present consciousness, perception and thought. An enlightened person must thus be said to have a quasi-ego, which operates as a focal point for quasi-memories, quasi-consciousness etc. A quasi-memory, as Parfit describes it, would be like a memory that is directly experienced, but because (in his examples) it has been artificially implanted into the mind, is not actually something that I experienced. Such a memory might, of course, be identified as *mine* and thus appear simply to be a false memory: but if we can imagine a directly-experienced memory which is not identified-with, this must be something similar to the memories of the enlightened. It is harder to imagine quasi-consciousness or quasi-perception, but perhaps we might get an inkling of these through the further extrapolation of Parfit's ideas.

4. Conclusion

Perhaps the outcome of this argument can be best summarised by using another version of Parfit's image of the coexistent person and series-person. The author of *Reasons and Persons* appears to have two coexistent identities as "Analytic Parfit" and "Buddhist Parfit", but, as with Parfit and Phoenix Parfit, the existence of the latter is not widely acknowledged. The book offers many hints of a recognition that key problems of philosophy like those of identity and ethics cannot be conclusively solved through an analytic approach and that reductionist utilitarianism is for him something more than an analytic position. Analytic Parfit, however, has not allowed Buddhist Parfit to follow through the full implications of this recognition or to recognise the contradictions it creates. Nevertheless, Buddhist Parfit has managed to leave a rich trail of useful insights which those more consistently Buddhist (or even those more consistently analytic) can use to aid the construction of a more adequate theory.

Bibliography

Adams, Robert Merrihew (1997), 'Should ethics be made more impersonal?' in Jonathan Dancy (ed.) *Reading Parfit*. Oxford : Blackwell

Ñanamoli and Bodhi (1995) (trans.) *Majjhima Nikaya* translated as *The Middle Length Discourses of the Buddha*. Boston: Wisdom Pubns.

Parfit, Derek (1984), *Reasons and Persons.* Oxford : Oxford University Press.

Williams, Paul (1989), *Mahayana Buddhism* London/New York: Routledge.

---

[1] My assertions about the Buddha's views in this paper ultimately come from a variety of sources. I shall not attempt to give full scholarly support to these assertions, since in any case their <u>historical</u> correctness is not relevant to the main purpose of the paper, which is one of philosophical argument. I shall therefore provide only one brief reference here: Ñanamoli and Bodhi (1995), Sutta 102. The views expressed in the whole of the paper are Buddhist in a broad philosophical sense, but not necessarily in a historical or scholarly sense.

[2] Although only the Cartesian version of the cognitive self is a completely immutable subject of knowledge, wherever there are taken to be determinate criteria of self-identity there must be some quantitative and immutable feature.

[3] The 'necessarily' here arises from a stipulation that an ego cannot be said to exist in a universe where there is no conflict with other egos.