

Double Troubles for Sidgwick's Dualism of Practical Reason

Ingmar Persson

Abstract

In *The Point of View of the Universe*, Katarzyna de Lazari-Radek and Peter Singer attempt to resolve Henry Sidgwick's 'Dualism of Practical Reason' between the rationality of egoism and the rationality of universal benevolence by undermining the former. I argue that, according to their interpretation at least, this dualism involves two troublesome steps: from egoism to impartiality and from impartiality to universal benevolence. I try to show that their attempt to undermine the rationality of egoism fails but go on to sketch another way of undermining it and establish an impartiality or universalizability that is arguably enshrined in the concept of morality. In contrast, the obstacles to the step from impartiality or universalizability to universal benevolence – to which de Lazari-Radek and Singer pay too scant attention – seem insurmountable. Questions about their hedonist conception of well-being are also raised.

1. Introduction

Katarzyna de Lazari-Radek and Peter Singer's *The Point of View of the Universe* (2014) is an impressively thorough exposition of hedonistic utilitarianism based on Henry Sidgwick's version of it. In the process of this exposition, they attempt to resolve his 'Dualism of Practical Reason' between the rationality of egoism and the rationality of universal benevolence by undermining the former.¹ I argue that, at least according to their interpretation, this dualism involves two troublesome steps: from egoism to impartiality and from impartiality to universal benevolence. In section 2 I try to show that their attempt to undermine the rationality of egoism fails but in section 3 I go on to sketch another way of undermining it and establish an impartiality or universalizability that is arguably enshrined in the concept of

¹ They advance essentially the same argument in some papers, e.g. 2017, but I shall refer mainly to the book.

morality. In contrast, the obstacles to the step from impartiality or universalizability to universal benevolence – to which de Lazari-Radek and Singer pay too scant attention – seem insurmountable, as will be contended in section 4, where their hedonist conception of well-being is also questioned.

2. Why the Evolutionary Explanation of Egoism is not Debunking

De Lazari-Radek and Singer propose to undercut the rationality of egoism – i.e. our special concern for our own well-being – by advancing a debunking evolutionary explanation of it. Since essentially the same evolutionary explanation works for kin altruism, our special concern for the well-being of members of our family and other relatives, they conclude that ‘partial reasons can be debunked’ (2014: 196) and that ‘all reasons for action are impartial’ (2014: 197). In other words, reference to particular individuals can be eliminated from our reasons for benevolence. By this means, they think that Sidgwick’s dualism can be ‘dissolved’ (2014: 197).

To begin with, when evolutionary explanations of attitudes are put forward, it is important to keep in mind the distinction between the reason or explanation *why* we have the attitude and *our* reason for having the attitude. For instance, the evolutionary reason or explanation why many of us have an aversion to having sex with our siblings may well be the risk of inbreeding such behaviour carries with it. This aversion may be widely spread among humans because those of our ancestors who exhibited it were likely to be more successful in having their genes replicated. But it was certainly not our ancestors’ reason for being averse to having sex with their siblings; they did not abstain from such sexual practices because they *regarded* them as increasing the risk of having genetically defective offspring. This consideration did not enter into the *content* of their reasons. They were averse to such sex simply because they were not sexually attracted by their siblings, not because of any consequences for offspring. The connection with the avoidance of deleterious consequences for offspring lies in the fact that *acting* in accordance with this aversion has a tendency to prevent such consequences, not in that the idea of this prevention is part of the content or object of the attitude.

If our reason for being averse to sexual intercourse with siblings was that it would put us at greater risk of having offspring with genetic abnormalities, we would not be averse to having intercourse with them if contraceptives were used, but this precaution will surely not make our aversion disappear. Nor will it disappear if we discover that we or they are sterile. This is because it is an aversion to having sex with siblings *per se*, regardless of its consequences for reproduction. Consequently, the availability of this

evolutionary explanation of our aversion to having sex with siblings is not debunking in the sense of showing that there is anything wrong with having this aversion *in itself*, which provides a reason to get rid of it. But since the link with eliminating the risk for inbreeding is found in acting in accordance with this aversion, it brings out that there is no reason to command abstention when sex between siblings occurs in circumstances in which there is no risk of conception.

The evolutionary explanation of why most of us have a strong desire to have sex with the opposite sex is obviously that this desire drives us to act in ways that promote genetic replication. Again, this is a normal consequence of acting on the desire; it is not a consideration that enters into the content of our sexual desire. This is a desire to have sex for the pleasure of it, without any thought of procreation. Our reason for having sex could of course be that it is a means of having children. Then we would discontinue this behaviour if we discovered that there is no chance of reproduction, but this is not what happens when we have sex in order to satisfy our sexual desire. Therefore, the proposed evolutionary explanation does not debunk this desire, does not show that there is anything wrong with this desire in itself.

The same holds for an evolutionary explanation of our special concern for our own well-being in terms of it in general being an effective means of promoting our reproductive fitness. This explanation cannot debunk our egoism or self-concern, for a reference to such matters as spreading our genes does not enter into the content of our self-concern just as it does not enter into the content of our sexual desire. The reference to this effect explains our being self-concerned because acting from this motive tends to have this effect. We are concerned about our own well-being for its own sake, not as a means of multiplying our offspring. If we discovered more effective means of procreation, we would be unlikely to give up the pursuit of our well-being in favour of them. Just like sexual satisfaction, our own well-being is intrinsically desirable for us, much more so than reproductive fitness – if the latter is intrinsically desirable at all.

It may be easier to comprehend that these evolutionary explanations are not debunking if we compare them to an evolutionary explanation of an attitude that has a good claim to be capable of debunking it. Consider a plausible evolutionary explanation of our *bias towards the near future* which modifies our self-concern, making us more concerned about our well-being when it is closer to the present. Thus, we are spontaneously disappointed when our pleasures are postponed and relieved if our pains are, though this makes no significant difference to the probability of their occurrence. So, our greater concern about our well-being when it is closer to the present is not a

concern motivated by its greater probability. It seems to be a greater concern about our well-being when it is temporally closer simply because it is temporally closer, for its own sake.

But it is a plausible hypothesis that evolution has equipped us with this bias because what is closer to the present is as a rule more probable and something that it is more urgent that we quickly consider dealing with. These are clearly matters of relevance for our reproductive fitness, but when we explicitly separate them from the temporal location of an event, we realize that the latter is in itself of no significance. As de Lazari-Radek and Singer maintain, following Sidgwick, ‘it is self-evident that a mere difference in time does not give some moments of our own existence greater significance than any other moments’ (2014: 191). Certainly, it is nevertheless difficult to overcome the bias towards the near because it is evolutionary ‘hard-wired’. Thus, we continue to give in to it, in spite of realizing that it is not being correlated with significant differences in probability, though we sense that this is irrational. Rationality requires that we focus our attention entirely on probability and possibilities of pertinent action. By contrast to sexual desire and self-concern, there is however nothing intrinsically desirable about temporal proximity. This is what makes the explanation of the bias towards the near debunking: it provides us with reason to abandon this bias in favour of an attitude with a content that we recognize as more sensible.

It follows that a *rational* egoism or self-concern must be cleansed of the bias towards the near – so as to become an interest in our own good that would be equally strong whatever the future timing of the good is – as opposed to our *spontaneous* egoism or self-concern which is under the influence of this bias. Such a more temporally neutral egoism clearly cannot be debunked by any appeals to evolutionary explanations.

Contrary to what has been argued here, de Lazari-Radek and Singer contend that ‘the principle of egoism is subject to a debunking evolutionary explanation’ (2017: 292; cf. 2014: 190–7). They reach the same verdict on kin altruism, and the evolutionary explanation is in both cases the familiar one of ‘promoting the survival of the genes we carry’ (2014: 194). They argue that ‘we have grounds for supporting the intuition for which there is no evolutionary explanation rather than one for which there is an evolutionary explanation’ (2014: 197). This argument strikes me as odd in view of the fact that they quote with approval Derek Parfit as asserting that the claim that ‘some attitude has an evolutionary explanation... is neutral. It neither supports nor undermines the claim that this attitude is justified’ (1987: 308). In contrast, their view rather suggests a presumption that evolutionary explanations in general are debunking.

De Lazari-Radek and Singer note: 'Denying the rationality of egoism leaves reason detached from our strongest sources of motivation, namely our desires to further our own interests and those of our family' (2014: 197 & 2017: 293). But I have argued that it is precisely our strong intrinsic interest in sex and our own well-being by contrast to temporal proximity which stands in the way of the explanation being debunking in the former case. They think that the proposed separation of reasons from motivation should not hold us back if we accept that 'normative reasons are independent of our present desires, wants, and beliefs' (2017: 293; 2014: 197–9). However, although this view of reasons denies any *conceptual* dependence of reasons on our motivation, it would still appear awkward if *in fact* it turns out that we are not much motivated to do what we have normative reasons to do. Is it not easier to believe in the existence of (conceptually) desire-independent normative reasons if they are such that evolution has designed us to be motivated by them than not to be motivated by them? Consequently, advocates of desire-independent reasons should not take lightly a *de facto* separation of reasons and motivation.

3. How the Rationality of Egoism Gives Way to Impartiality

The upshot of the preceding section is that I cannot see how de Lazari-Radek and Singer can justify their claim that the evolutionary explanations of egoism and kin altruism are debunking. This raises the question whether there is any other way to undermine the rationality of these attitudes. An implication of my reasoning above would seem to be that the source of their irrationality must be found in their content. Parfit undertakes such a project when in *Reasons and Persons* (1987: pt. 3) he argues that personal identity is not what rationally matters on the basis of an analysis of what it consists in.

He analyses our identity over time as consisting in psychological continuity when it 'has the right kind of cause' and when it 'has not taken a "branching" form' (1987: 207). Psychological continuity is composed of 'the holding of overlapping chains of strong connectedness' (1987: 206). There is strong psychological connectedness if the number of psychological connections 'over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person' (1987: 206). Examples of psychological connections are having memories of the past experiences of someone, having intentions that are carried out by someone in the future (1987: 205) and having the same traits of character as someone at another time in the past or future (1987: 207). The right kind of cause can be the normal cause consisting in the connections being underpinned by identical

brain parts. It is however the non-branching constraint which is crucial in his argument that personal identity is not what matters.

Parfit presents a thought-experiment in which your consciousness and mind are *divided* (1987: ch. 12). Imagine that each hemisphere of your brain is capable of maintaining the whole of your mind and, thus, all the psychological connections that obtain between one day and another in your life. Imagine further that the neural fibres, the *corpus callosum*, between the hemispheres are cut and the two hemispheres are taken out of your skull and transplanted into two different, brainless bodies, each of which may be assumed to be a perfect replica of your body. Both transplantations are successful and, as a result, there is a consciousness and mind like yours in each of the bodies, with your memories, character traits, and so on.

Parfit contends that in this case you are not identical to anyone of the two persons embodied in the bodies into which your brain-halves have been transplanted because your consciousness and mind have branched or divided. If your undivided brain had been transplanted into a single body, or one of the two transplants had failed, you would have been identical to the resulting person, but not if both succeed. For we cannot identify you with each of the resulting persons, since they are evidently distinct from each other, and it would be arbitrary to identify you with one of them. Therefore, you are not identical to anyone of them; you have ceased to exist and been succeeded by two numerically distinct persons who are virtually indistinguishable from you in macroscopic respects. But, Parfit maintains, this is no less *good* for you than it would have been if there had been only one successful transplant, and you would have been identical to the post-operative person. Consequently, it does not matter that identity fails to be preserved.

What matters instead of identity, Parfit suggests, is 'psychological connectedness and/or psychological continuity with the right kind of cause' (1987: 262). It does not matter whether or not it branches, though branching rules out personal identity. He goes on to claim that 'the right kind of cause could be any cause' (1987: 262). He fantasizes about being teletransported: 'The Scanner here on Earth will destroy my brain and body, while recording the exact states of all my cells. It will then transmit this information by radio' to 'the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine' (1987: 199). Although the psychological connections between Parfit on Earth and his replica on Mars do not have the normal cause in the shape of the persistence of one and the same brain, he proposes that being teletransported is as good as ordinary survival (1987: 285–7).

But having gone this far, we might wonder: does it matter whether there is *any* causal link between Parfit on Earth and his replica on Mars? Suppose that the radio signals never reach the Replicator but that nonetheless it inexplicably creates a replica of Parfit. It would seem that this is as good as being teletransported (as I argue in most detail in 2005: ch. 23). This would be an important conclusion for it would mean that, under the pressure of rational reflection, the partiality of our egoistic concern for our own well-being has been transformed into an *impartial* concern for the well-being of a certain *kind* of individual. If our egoistic concern for our own well-being is merely transformed into concern for the well-being of individuals who are psychologically continuous and connected to *us*, it is still a partial or biased concern in virtue of the reference to us. But if this reference is dropped and the concern is for the well-being of a certain kind of individual regardless of whether it has originated from us, our concern for it can be impartial, by being based only on features of this individual that are universal in the sense of involving no reference to particulars.

Such an impartial concern is *universalizable* in the sense that Richard Hare claimed is characteristic of moral judgements, 'namely that they entail identical judgements about all cases identical in their universal properties' (1981: 108). Suppose that I am inclined to securing a smaller benefit to myself rather than letting another person who is worse off than me get a bigger benefit because, as opposed to the other person, I have a universal feature F. Then my inclination is universalizable in the sense required by morality only if I am prepared to accept that in a hypothetical situation in which I lack F and the other person possesses it – this switch must be conceivable since the feature contains no reference to particulars – the other person receives the bigger benefit. It is quite likely that I am not prepared to accept this. Hare points to the resemblance between his requirement of universalizability and the Golden Rule of Christianity. De Lazari-Radek and Singer fill in that other world religions have acknowledged principles similar to the Golden Rule as 'the essence of morality' (2014: 193).

4. Why Universal Benevolence Cannot be Derived from Impartiality

Let us proceed on the assumption that there is a moral requirement of impartiality or universalizability and ask whether it could move us to universal benevolence. Hare claimed that it could because "'I" is not a wholly descriptive word but in part prescriptive' (1981: 96), that is

when I say that somebody who would be in a certain situation would be *myself*, in so saying I express a concern for that person in that hypothetical situation which is normally greater than I feel for *other people* in the same situation. To recognize that that person would be myself is already to be prescribing that, other things being equal, the preferences and prescriptions of that person should be satisfied. (1981: 221)

Unfortunately, this idea is incompatible with Parfit's claim that personal identity is not what matters (cf. Persson, 2005: 347–8). According to Parfit, the fact that somebody is identical to me is no reason in itself to be concerned that his preferences be fulfilled. The prescriptivity of 'I' is also implausible for reasons that Hare himself expressed earlier in his career: he then considered the possibility of someone who 'sticks to his judgements even when they conflict with his own interests in hypothetical cases' (1963: 162). His example was that of a Nazi who imagines being a Jew. But aims for the sake of which we could be ready to sacrifice our long-term well-being need not be obnoxious; they could be admirable artistic, scientific, aesthetic or athletic ideals. It is hard to believe that it would be irrational to sacrifice one's long-term well-being in the name of such ideals, especially if 'personal identity is not what matters' (as I argue in 2005: pt. IV). Such idealism could well be universalizable.

Speaking of the Golden Rule and similar ideas in other world religions, de Lazari-Radek and Singer concede: 'Admittedly, these rules do not require us to adopt universal benevolence, but they do require impartiality' (2014: 193). They appear to voice an objection to Hare somewhat akin to mine when they write 'moral language appears to allow me to commit myself to an ideal that is independent of my own desires', for which 'I am willing to sacrifice everything that is dearest to me' (2014: 124).² But, for reasons that escape me, they apparently affirm that their 'debunking' of egoism may carry them not merely to impartiality but to universal benevolence when they claim: 'If the rationality of egoism can... be put in doubt, we can tentatively conclude that all reasons for action are impartial, and the dualism [that Sidgwick feared] can, at least at the level of rationality, be dissolved' (2014: 197). If this is right, there must be more to their debunking than I can find.

Because universal benevolence is not 'the direct outcome of any evolutionary process' or 'any other obvious, non-truth-tracking' process, de Lazari-Radek and Singer conjecture: 'Like our ability to do higher mathe-

² Their objection is however in some respects different from mine, since they interpret his position differently, in particular they miss his idea of the prescriptivity of 'I'.

matics, it can most plausibly be explained as the outcome of our capacity to reason' (2014: 193). But, as they concede, this 'does not prove that it is a substantive normative truth' (2014: 193). Reason can unfortunately lead us astray, and it is often controversial when it does. For instance, there is not more of an evolutionary explanation of the common anthropocentric idea that the well-being of all humans is of an equal value that is higher than that of all non-human animals than there is of universal benevolence (which includes the well-being of non-human animals as well). The former is rather a product of 'reason' or 'culture' like the latter. De Lazari-Radek and Singer would surely reject the anthropocentric view, but if they were to do so on the ground that it is due to a 'non-truth-tracking' process, they could be charged with begging the question against its truth. So, we need a better argument in favour of universal benevolence than this.

Is there then any way of justifying the derivation of universal benevolence by means of a requirement of universalizability? De Lazari-Radek and Singer endorse Parfit's theory of reasons, so they might propose that we have an objective, desire-independent overriding reason to desire for its own sake well-being in proportion to its amount and irrespective of who its recipients may be. This could make it irrational to desire to sacrifice well-being, whether one's own or that of others, in the name of some ideal.

Why believe that there are such objective reasons? One argument Parfit puts forward is that there are desires that are definitely or indisputably irrational but that cannot be declared to be irrational unless there are such reasons. They are intrinsically irrational. De Lazari-Radek and Singer appeal to a well-known example of such an attitude which he calls 'Future Tuesday Indifference' (1987: 124; see also 2011: vol. 1, 56.). 'Throughout every Tuesday', Parfit writes, a man who has this strange attitude

cares in the normal way about what is happening to him. But he never cares about possible pains or pleasures on a *future* Tuesday. Thus he would choose a painful operation on the following Tuesday rather than a much less painful operation on the following Wednesday. (1987: 124)

This Future Tuesday Indifference 'does not depend on ignorance or false beliefs' (2011: vol. 1, 56) about pain, personal identity, or time. Therefore, it would not be undermined by the 'ideal deliberation' required by a *deliberative theory* of practical rationality, that is, it would pass 'if he knew the relevant facts, was thinking clearly, and was free from distorting

influences' (1987: 118). Here 'facts' are *natural* facts, i.e. facts of the kind investigated by natural and social sciences.

But, I want to argue, a couple of scenarios bring out that such an attitude could not survive ideal deliberation. Suppose that late one evening this man learns that he must now choose between having an intense pain in a few minutes and a much milder pain on the day following the day of the earlier pain. The problem is that he does not know whether it is just before Monday midnight, but that it will be past midnight and Tuesday in a few minutes when he will feel the intense pain, or if it is already Tuesday. If the pain will occur on the other side of midnight, it will be a Future Tuesday pain for him which he should now prefer to the milder pain on the following day. However, if it is already Tuesday, the intense pain is a Present Tuesday pain for which he should care 'in the normal way'. Then he should now prefer to have the milder pain on the following day. But clearly, it cannot reasonably be held to matter whether the intense pain that he will feel in a few minutes will be a Future Tuesday pain or a Present Tuesday pain, what its position is in 'a conventional calendar' (1987: 124). What matters is simply that it will be intense and will be felt by him in a few minutes.

A variation of this scenario underscores the point. Consider an intense pain which begins on a Monday some time just before midnight, but extends into the following Tuesday, say, it begins five minutes before midnight and ends five minutes after. Since the man 'cares in the normal way about what is happening to him' except on future Tuesdays, on Monday evening he will fear and wish to prevent the five minutes of pain that he will have that day but be indifferent towards the five minutes of this pain that will follow after midnight, on what is now still a future Tuesday. When he starts feeling the pain, he will dislike and wish to stop the stretch of the pain that will happen before midnight but not care about being exposed to the five minutes of it occurring after midnight. This is so, though he is aware that a few minutes later, he will dislike and wish to stop this stretch as well. There is surely an inconsistency in not wishing to stop part of an unchanging pain that you realize that you will soon wish to stop, which means that it will not survive ideal deliberation.

It is also noteworthy that if the pain had started five minutes earlier and occurred in its entirety before midnight, the man would have wished to stop all of it. He is committed to holding that merely in virtue of the fact that a pain starts five minutes later, a part of it cannot be intrinsically worse for him. And if he does not know exactly when it is midnight, he does not know how much of the pain he should be indifferent to. But for a man who 'knows that Tuesday is merely part of a conventional calendar' (1987: 124), it is

clearly irrational to let the extent to which a pain occurs on a Tuesday determine his attitude to it in this way.

Parfit describes his man's indifference as 'a bare fact' (1987: 124). This may suggest that he regards it as akin to the dislike of pain or the liking of pleasure, but he rightly maintains that these attitudes cannot be 'either rational or irrational': 'Whether we like, dislike, or are indifferent to these various sensations, we are not responding or failing to respond to any reasons' (2011: vol. 1, 53). If the man's indifference to pain on future Tuesdays were like our dislike of pain, it would indeed be true that ideal deliberation could not remove it, but it would also be true that it did not qualify as irrational. Since this is not what it could mean that his indifference is 'a bare fact', what could it mean? Not having an answer, my conclusion is that either ideal deliberation on the examples I have provided will show Future Tuesday Indifference to be irrational or it is non-rational. It seems to me that if somebody has this attitude in these examples, it *must* be true either that their deliberation has not been ideal or that the attitude is non-rational, insensitive to reasons – unless it is 'so bizarre, we cannot usefully discuss whether it is irrational' (1987: 124).

However, we should also take a quick look at the nature of these putative desire-independent reasons (for more discussion of this issue, see Persson, 2013: ch. 12). Parfit maintains, for instance, that we should be averse to pain for its own sake: 'We all have a reason to want to avoid, and to try to avoid, all future agony' (2011: vol. 1, 76). What could the *propositional content* of these reasons be? A reason for an attitude must have a content in virtue of which it 'counts in favour of' (2011: vol. 1, 31) the attitude; for example, the content of the reason to refrain from taking a toxic pill is that it will cause something that you are determined to try avoiding, namely your death or ill health. Since we want to avoid agony for its own sake, the reason-providing fact here cannot be about anything beyond agony itself, but it cannot simply be the tautology that agony is agony.

An answer that readily suggests itself is that agony, and pain in general, is in itself *bad*. Now Parfit claims: 'It is our hedonic likings and dislikings... that make these conscious states good or bad' (2011: vol. 1, 3; 55). He also claims that 'the most important uses of "good" and "bad" are... reason-implicating' (2011: vol. 1, 39). This means that if this is the sense in which pain is said to be intrinsically bad, pain is said to have some unspecified property which gives us a reason to be averse to pain for its own sake. Since it is 'our hedonic likings and dislikings... that make these conscious states good or bad', it must be these likings and dislikings that provide reasons as well.

Does the intrinsic badness of agony then simply consist in it being disliked for its own sake? No, for although Parfit holds that in ‘any possible world, pain would be in itself bad’ (2011: vol. 2, 489), and he claims: ‘When we are in pain, what is bad is not our sensation but our conscious state of having a sensation that we dislike’ (2011: vol. 1, 2), he insists that the badness of pain ‘does not follow from the meaning of the words “bad” and “pain”’ (2011: vol. 2, 490). His rejection of naturalism implies that ‘no such normative truths could be analytic’ (2011: vol. 2, 490). If the intrinsic badness of pain were entailed by it being intrinsically disliked, we would have a naturalist account of its intrinsic badness. Parfit also holds that we know a priori, ‘merely by thinking’ (2011: vol. 2, 490), and independently of experience, that pain is in itself bad in any possible world in which it is disliked. But how can we know this a priori unless the intrinsic badness of pain is entailed by its being intrinsically disliked which would imply a naturalist account of intrinsic badness? This is a question Parfit does not seem to answer.

As mentioned, de Lazari-Radek and Singer compare our capacity to engage in fundamental moral reasoning to our ability to do higher mathematics, which provides us with knowledge a priori. But the elements of the method of establishing the reliability of moral intuitions they specify – namely ‘independent agreement of other careful thinkers’ and the absence of explanations like evolutionary ones (2014: 195)³ – have very little resemblance to mathematical proofs.

A further source of worry is the notion of well-being, or what it is that makes life go well. I assume that well-being is the object of benevolence, and that universal benevolence is something like the same concern for the same amount of well-being whoever has it. This understanding of the object of benevolence is implied when de Lazari-Radek and Singer write: ‘The principle of universal benevolence needs a theory of well-being, or else it is empty of content’ (2014: 196). Therefore, the claim that we have an objective reason to desire well-being would be an empty or vacuous claim without a theory of well-being. But then it could be a problem that it is controversial what well-being consists in. True, this need not be a problem if reasons were desire-dependent, for then we could allow that benevolent people, who have different desires about what to promote because their conceptions of well-being differ, could have reasons to pursue different things for beings, themselves and others. But this will not do if such reasons are desire-independent.

³ See also their statement of Sidgwick’s conditions of self-evidence (2014: 94–5).

De Lazari-Radek and Singer come down in favour of a version of hedonism which declares that 'what is ultimately good is pleasure, understood as desirable consciousness' (2014: 252). Now a much-discussed objection to hedonist theories is generated by Robert Nozick's 'experience machine'

that would give you any experience you desired. Superduper neurologists could stimulate your brain so that you would think and feel you were writing a great novel, making a friend, or reading a book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life? (1974: 42–3)

Nozick thinks the answer to this question is a resounding 'no' because we want to be in contact with a reality that exists independently of our experiences.

The thought-experiment of the experience machine has however various irrelevant features. Let us peel off these and focus on the following simple claim, in de Lazari-Radek and Singer's words: if 'the lives of P and Q are, from the inside, indistinguishable, the hedonist must say that P and Q have the same level of well-being' (2014: 259), provided that they differ only in that P's experiences correspond to an independent reality. More concretely, suppose that P and Q lead lives which contain the very same experiences of being engaged with a lot of people in various ways. In P's case these experiences correspond to actual engagements with flesh-blood people, whereas in Q's case solipsism is true, and Q is the only conscious being. De Lazari-Radek and Singer end up saying that they 'do not claim to have shown that' the judgement that 'P's life was better for her than Q's life was for her' 'would be an unreasonable judgment' but that hedonist theories 'are still viable' (2014: 261).

I believe, however, that it is incontestable that human beings are in general intensely social creatures. Virtually all sane people strongly want there to be other human beings around, and the thought of living a solipsistic world would make them feel horribly lonely. If there are no independently existing bodies of which they have sense-impressions, there is nothing apart from themselves that can possess minds. It is as hopeless to find a debunking evolutionary explanation of our desire to be in touch – literally! – with other conscious beings as to find such explanations of our sexual desire, egoism or kin altruism. On the other hand, my argument with respect to Future Tuesday Indifference prevents me from condemning an indifference to being

surrounded by other people as intrinsically irrational in Parfit's sense. But, although I think it is evident that a very important part of our well-being consists in something that is felt or experiential, I find it very hard to deny that for an overwhelming majority of us it also includes something trans-experiential, like the satisfaction of a desire that one's friends do not successfully deceive one behind one's back, or that a project one has spent the better part of one's life working on is successfully completed after one's death.

Imagine, however, that hedonism is 'still viable'; then de Lazari-Radek and Singer face another problem, already hinted at. If there are objective, desire-independent reasons to promote everyone's well-being, it would seem that well-being would have to mean the same for all and sundry whilst, to repeat, if these reasons are desire-dependent, the well-being they give benevolent people reason to promote could depend on whether they favour, say, a purely experiential or a trans-experiential conception of it. Our grasp on what objective reasons there are is regrettably not so firm that we could settle what well-being comes to by consulting them. The reasoning typically goes in the opposite direction from intuitions about what desires seem indisputably rationally required or impermissible to what objective reasons there must be.

To sum up. I have tried to show that Sidgwick's dualism of practical reason, as interpreted by de Lazari-Radek and Singer, involves two troublesome steps: one step from the partiality of egoism (and kin altruism) to impartiality (or universalizability) and another step from impartiality to universal benevolence. I have rejected their attempt to take the first step of denying the rationality of egoism by means of appealing to a debunking evolutionary explanation of it, but sketched what I think is a feasible way of taking this step. However, I cannot envisage any tenable way of taking the second step. This means that I cannot see how we can be rationally required to be universally benevolent, but of course this does not imply that it is rationally impermissible to adopt this attitude. The further move to a hedonist construal of the object of benevolence is also doubtful.

It does seem to me plausible to hypothesize that an idea of impartiality is part of 'the essence of morality'. In order for a community to function harmoniously for the benefit of its members, they have to agree on some set of norms of behaviour, and to reach such an agreement, they would have to rise above their spontaneous partiality. In the more distant past when societies were small and largely isolated from each other, and the views of their members were consequently more homogeneous, it must have been a lot easier to establish workable agreements than it is in the globalized world of

today with societies with millions and even billions of citizens. My interpretation of impartiality is not meant to capture a notion that has in fact been in operation in these societies but to be a viable refinement of it.

References

- De Lazari-Radek, K. & Singer, P. (2014). *The Point of View of the Universe*. Oxford: Oxford University Press.
- De Lazari-Radek, K. & Singer, P. (2017). Parfit on Objectivity and the 'Profoundest Problem of Ethics'. In P. Singer (ed.): *Does Anything Really Matter?*, 279–96. Oxford: Oxford University Press.
- Hare, R. M. (1963). *Freedom and Reason*. Oxford: Clarendon Press.
- Hare, R. M. (1981). *Moral Thinking*. Oxford: Clarendon Press.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Parfit, D. (1987). *Reasons and Persons*, reprinted with corrections. Oxford: Clarendon Press.
- Parfit, D. (2011). *On What Matters*. vols. 1 & 2. Oxford: Oxford University Press.
- Persson, I. (2005). *The Retreat of Reason*. Oxford: Clarendon Press.
- Persson, I. (2013). *From Morality to the End of Reason*. Oxford: Oxford University Press.

Ingmar Persson
Ingmar.persson@filosofi.gu.se
ingmar.persson@philosophy.ox.ac.uk