

What's the Point of Self-consciousness? A Critique of Singer's Arguments against Killing (Human or Non-human) Self-conscious Animals

FEDERICO ZUOLO

University of Hamburg

Singer has argued against the permissibility of killing people (and certain animals) on the grounds of the distinction between conscious and self-conscious animals. Unlike conscious animals, which can be replaced without a loss of overall welfare, there can be no substitution for self-conscious animals. In this article, I show that Singer's argument is untenable, in the cases both of the preference-based account of utilitarianism and of objective hedonism, to which he has recently turned. In the first case, Singer cannot theoretically exclude that a self-conscious being's stronger preferences may only be satisfied by killing another self-conscious being. In the second case, he fails to demonstrate that the rules of ordinary morality, demanding that killing be strictly forbidden, could not frequently be overruled by the principles of esoteric morality. In both cases, his theory cannot solve the classical utilitarian problem of prohibiting the killing of people in secret.

I. INTRODUCTION: UTILITARIANISM AND SELF-CONSCIOUSNESS

Act-utilitarianism has traditionally had difficulty justifying the prohibition on the killing of individuals. Diverse strategies have been devised to prevent the idea that killing people could be justified. It has been said that killing is directly wrong in so far as it deprives the world of that individual's future enjoyable experiences, and indirectly wrong because it creates anxiety in others who fear the possibility of this fate. It is also thought to be unlikely that, in practice, utilitarian calculus could ever really justify the killing of an individual. However, these counterarguments do not represent a convincing response to the theoretical and practical admissibility of killing an individual under certain conditions (without prior notice, without pain and public knowledge thereof). Even if we curtail the probability of this event, act-utilitarianism still faces a problem that makes it unacceptable to many people. This is one of the reasons why some theorists have switched to rule-utilitarianism or have preferred a deontological approach. But in this article I am concerned only with the problem of killing in act-utilitarianism. Probably the last and most challenging act-utilitarian

argument against the permissibility of killing people is that provided by Peter Singer.

Famously, to assess the moral permissibility of killing a being, Singer distinguishes between merely conscious beings and self-conscious beings. The killing of merely conscious beings is wrong only in so far as it precludes future enjoyable experiences. But conscious beings may be replaced by bringing new, similar beings into life. This is not permissible for self-conscious beings, which are not replaceable qua biographically valuable and individually distinguished beings. In this article, I want to challenge the overall tenability of Singer's argument against the killing of self-conscious beings. After explaining the nature of preference utilitarianism in [section II](#), in [section III](#) I focus on the preference-based argument developed in *Practical Ethics*, because it seems to me to be the most refined and powerful argument against the permissibility of killing people. In [sections IV](#) and [V](#), I lay out some considerations regarding the revision of Singer's view in his more recent *The Point of View of the Universe*.¹ My point will be that in both cases, Singer's position on this issue is troublesome. This is because in *Practical Ethics* he cannot defend, as he is committed to doing, the prohibition on killing self-conscious animals in principle, while from what he says in *The Point of View of the Universe* we can doubt whether he can actually do so in practice. Most of the article will be concerned with the preference-based argument, because it seems like the most suitable and original strategy within act-utilitarianism to justify the prohibition on killing most animals (including human beings). However, I will conclude with a critique of the idea of esoteric morality, because such an idea justifies the ordinary prohibition on killing both in *Practical Ethics* and in *The Point of View of the Universe*. Although Singer at least partially revises his position in his latest work, we can still carefully analyse the preference-based argument against killing animals. Indeed, the preference-based approach is still quite a popular position in philosophy and economics. Moreover, Singer's recent co-authored work is an overall defence of Sidgwick's perspective. Hence, his latest work should not be seen as a full abandonment of his

¹ I will focus in particular on the second edition of P. Singer, *Practical Ethics* (Cambridge, 1993), rather than on the third edition (2011), because the second edition represents the most coherent attempt to develop a preference-based argument, whereas the latest version is a mix of a preference-based argument and hedonistic argument. This latter approach has been fully defended in K. de Lazari-Radek and P. Singer, *The Point of View of the Universe* (Oxford, 2014). For a reconstruction of the ambiguity of Singer's arguments in the third edition of *Practical Ethics*, see S. Kagan, 'Singer on Killing Animals', *The Ethics of Killing Animals*, ed. T. Višak and R. Garner (Oxford, 2015), pp. 136–53.

previous position.² In sum, while I will follow Singer's thought up to his latest theory, the core of my argument will address the plausibility of a utilitarian preference-based argument against the permissibility of killing, whose best formulation can indeed be found in Singer's work.

Although valid for both animals and non-animals, the following argument will focus in particular on self-conscious human beings so as to circumscribe any doubts concerning Singer's capacity to justify the prohibition on killing in a way that is most favourable to his argument. Indeed, common-sense morality has a strong bias in favour of humans. But if we can prove that Singer's argument against the killing of self-conscious animals (in particular people) is unconvincing in so far as it is insufficiently strong to show the moral wrongness of killing persons, the same will also hold with respect to non-human self-conscious animals.

I intend to reconstruct and cast doubt on Singer's view. This critical enterprise can help test the convincingness of a relatively overlooked aspect of Singer's theory, concerning the normative tenability of his view about self-conscious beings and persons in particular. Singer's other positions challenging common sense on the treatment of animals, elders, new-born infants, severely mentally disabled human beings, and embryos, have already been widely discussed, while less attention has been paid to the tenability of his view with respect to issues of the life and death of ordinary human beings. In sum, I will mostly focus on the tenability of Singer's view regarding the killing of self-conscious animals, without directly discussing two related issues, namely his thesis on killing merely conscious animals and the empirical reliability of his distinction between conscious and self-conscious animals.

A final clarification is in order before we begin. As stated, in this article I will try to assess Singer's position. To do so, I will accept all his assumptions and check their tenability. This means that I will not take into account attempts to defend utilitarianism from the unwelcome implication of replaceability, as recently put forward by Tatjana Višak. In particular, I will accept Singer's assumption that we can compare the lives of existing beings with not-yet-existing or contingent beings, which is rebutted by Višak.³

² I am grateful to an anonymous reviewer for pressing me to clarify this point.

³ T. Višak, 'Do Utilitarians Need to Accept the Replaceability Argument?', *The Ethics of Killing Animals*, ed. Višak and Garner, pp. 117–35. For a critique of Višak, see N. Holtug, 'The Value of Coming into Existence', *The Ethics of Killing Animals*, ed. Višak and Garner, pp. 101–14.

II. THE CONTOURS OF PREFERENCE UTILITARIANISM

As is well known, in *Practical Ethics* Singer subscribes to preference utilitarianism:

According to preference utilitarianism, an action contrary to the preference of any being is, unless this preference is outweighed by contrary preferences, wrong. Killing a person who prefers to continue living is therefore wrong, other things being equal. That the victims are not around after the act to lament the fact that their preferences have been disregarded is irrelevant. The wrong is done when the preference is thwarted.⁴

The reason for this formulation of utilitarianism seems to be that classical utilitarianism did not provide sufficient grounds for prohibiting the killing of a being. As known, the objection is as follows: If the life of an individual is taken away while sleeping, without suffering or notice, no wrong is committed provided that the killing of such an individual is counterbalanced by more weighty considerations in terms of aggregate welfare. More specifically, this seems to justify the overall practice of rearing animals for food production provided that the life of the animal is worth living. If such animals are *replaced* after being killed we seem to have no direct argument against killing them because we compensate for the lack of future enjoyable experiences for an animal with equivalent experiences contained in the life of another equivalent animal. If this argument tells us why it is permissible to rear, kill and replace animals, the classical utilitarian reason for meat-eating derives from the pleasure humans receive from eating meat.

As is well known, Singer's strategy for rejecting this implication consists in drawing a distinction between conscious and self-conscious animals. Merely conscious animals experience pleasure and pain and are therefore included in the utilitarian calculus, but do not have subjective experiences extended from the past to the future. By contrast, self-conscious animals, besides the mere level of consciousness, have complex expectations regarding the future, memory of the past, and an irreducible individuality that we cannot replace. 'So perhaps the capacity to see oneself as existing over time, and thus to aspire to longer life (as well as to have other non-momentary, future-directed interests) is a characteristic that marks out those beings who cannot be considered replaceable.'⁵ Such a future-oriented characteristic is particularly prominent in persons and to a

⁴ Singer, *Practical Ethics*, p. 94.

⁵ Singer, *Practical Ethics*, p. 125.

lesser extent in some animals (mainly mammals).⁶ '[P]ersons are highly future-oriented in their preferences. To kill a person is, normally, to violate not just one, but a wide range of the most central and significant preferences a being can have.'⁷

With these clarifications made, we might ask: But why is killing wrong? What is so special about thwarting preferences? To answer these questions we must dig deeper into Singer's theory of value. In what follows I will analyse two alternative theories regarding the sources of value for individuals. The first is the satisfaction-based theory of value and the second is the object-dependent theory of value.⁸ I will argue that the first option, supported by Singer for a long time, cannot justify the prohibition on killing self-conscious animals. The second option is endorsed by Singer in the recent reformulation of his view. I will argue that the first fails to justify theoretically the prohibition of killing persons. The second is likely to be problematic in practice because of its dependence on esoteric morality.

III. THE SATISFACTION-BASED THEORY OF VALUE

According to the satisfaction-based theory of value, it is the satisfaction of a preference that confers value upon a certain state of affairs. This account is in principle content-independent, because what counts is the fact that a preference is satisfied irrespective of the type of preference that is satisfied or not satisfied. To better appreciate this account we must, first, understand what it is that determines the weight of one's preference, given its content independency. We can think that the weight ultimately depends on the *intensity* and *permanence over time* of a preference. This in particular gives justice to the idea that taking the life of a self-conscious being is a serious wrong because it frustrates complex and long-standing preferences regarding the future. Accordingly, if we consider the intensity and duration involved in the life of a self-conscious being, we may have sufficiently strong reasons to justify the prohibition on killing.

Before proceeding we must inquire a little further into what the preference for continuing one's life is. Singer's use of the idea of preference for continuing one's life seems to include three distinct preferences: the *direct preference* for the continuation of one's life, the

⁶ Singer is not clear on where we ought to draw the line. He says that for precautionary reasons we could at least include chickens but probably not fish, thus ruling out the possibility of rearing most animals typically eaten by humans.

⁷ Singer, *Practical Ethics*, p. 95.

⁸ I draw on W. Rabinowicz and J. Österberg, 'Value Based on Preferences: On Two Interpretations of Preference Utilitarianism', *Economics and Philosophy* 12 (1996), pp. 1–27, for the formulation of the first two alternatives with slight modifications.

indirect preference for the continuation of one's life, and the *overall preference* for the continuation of one's life.⁹ In the direct preference for the continuation of one's life, one's self-conscious preference for continuing to exist is *simpliciter* the direct object of one's preference. In the indirect preference for the continuation of one's life, one desires to continue one's life as a means to pursue and achieve a number of other future-oriented desires and plans. Here, the continuation of one's existence stands in an instrumental relation to the pursuit of the other direct object of concern. The overall preference for the continuation of one's life is the resulting combination of the direct and indirect preference for continuing one's life. Why are such distinctions important? They are interesting because they allow us to better identify which kind of preference is being thwarted in the act of killing. While there is a sort of general transitive relation between these preferences – it would be utterly inconsistent to have preferences regarding one's future, thus an indirect preference for continuing one's life, without also having a direct preference thereof – this is not biunivocally so. Indeed, having a direct preference for continuing one's life does not per se also imply having correspondent indirect preferences, because these latter preferences depend on one's specific plans and future desires that are not per se dependent on one's direct preference. Moreover, the direct preference may be conditional upon the fulfilment of certain other preferences. For instance, a political activist fighting for the liberation of her country from the oppression of a dictator may only wish to continue her life if she has a reasonable expectation of seeing the dictatorship overturned.

These distinctions help clarify the fundamental components of the prohibition on killing self-conscious animals. However, they seem to have some troubling implications. First, they justify hugely different moral assessments of the killing of diverse people. Second, in principle they allow any frustration of one's preferences, even the impossible ones, to be considered morally wrong. Third, there is no fully convincing argument that rules out the permissibility of replacing self-conscious beings. Singer is certainly aware of all this; however, I contend, drawing out their full implications may have more troubling and far-reaching consequences than expected.

III.1. *Unequal disvalue of killing?*

The first point above regards whether the killing of different persons is equally wrong. As is well known, Singer does not subscribe to the idea

⁹ For the distinction between the first two types of preferences, I elaborate on the hints sketched by T. Višák, *Killing Happy Animals: Explorations in Utilitarian Ethics* (Houndsmill, 2013), p. 37.

that some beings (for instance, persons) have an intrinsic property or set of properties that ground the ascription of an equal moral status. In so far as it is based on the idea of dignity as an intrinsic value, the traditional religiously inspired or Kantian morality cannot measure the wrongness of the killing of a person, because persons have incomparable value. On Singer's account, by contrast, only the weight of interests matters. But, interests notoriously vary, not only among species but also within a species.

If the wrongness of killing depends on the frustration of one's preference for continuing to live, we may say that we ought to weigh such a preference in order to establish to what degree the killing of an individual is wrong. This is necessary because both direct and indirect preferences to continue to live vary depending on a number of idiosyncratic factors and are different from individual to individual. If the importance of such preferences depends on one's plans for the future, one's commitments, and one's complex volitions, it follows that the wrongness of killing also depends on such future-oriented preferences. Singer admits this characteristic of his account by mentioning a non-problematic case: "Towards the end of life, when most things that might have been achieved have either been done, or are now unlikely to be accomplished, the loss of life may again be less of tragedy than it would have been at an earlier stage of life."¹⁰ But comparing the disvalue of the death of a person in two different stages of life does not confront the real problem that lies behind this account, or at least does not show it as vividly as it should.

Consider two very ordinary (human) persons. One (α) is a very proactive individual who engages in a number of activities, is very optimistic, educated, and has a host of plans for her future (having a family, pursuing a good career, helping others, etc.). The other person (β) is a very lazy or depressed individual, with very few plans for her future, living day to day and expecting nothing from life. (α) has both direct and indirect strong preferences for continuing to live; (β) has very weak indirect preferences and her overall preference relies mostly on the direct preference. If we had to assess the wrongness of killing these persons we should conclude that the killing of (α) is a serious wrong, while the killing of (β) is a much less serious wrong, if it is a wrong at all. And we might have a reason to kill (β) if (β)'s existence thwarts (α)'s existence. Or to put things in a slightly less cruel way, we could say that if we were in a situation in which both (α) and (β) were in danger and we could only save one of them, we would have strong reason to save (α). Hence we come up with a very revisionary normative conclusion holding

¹⁰ Singer, *Practical Ethics*, p. 130.

that there are huge variations in the moral assessment of killing (or not saving a person) depending on the type of interest that a person has in continuing to live. Such inequalitarian implications also concern the assessment of the value of individual lives.

So it would not necessarily be speciesist to rank the value of different lives in some hierarchical ordering.... In general it does seem that the more highly developed the conscious life of the being, the greater the degree of self-awareness and rationality and the broader the range of possible experiences, the more one would prefer that kind of life.¹¹

That such an implication should also be drawn for differences among human persons, for Singer, should come as no surprise. More generally, that Singer's normative implications are deeply revisionary of standard morality is not a daunting problem for him, as he has already coherently challenged a number of deeply entrenched moral intuitions regarding embryos, infants, elders, and so on. However, this case is not so easy, as we will see below.

In the light of this, one may ask whether an individual's overall preference for continuing to live is sufficiently weighty to generate a duty not to kill that being. The simple answer is that it depends. On a satisfaction-based account, preferences vary and we have no grounds to ensure that they are weighty enough and that such weight is more or less equal across persons. The assessment of the weight of a person's interest in continuing to live must proceed alongside an assessment of the relative weight of others' interests, because from a utilitarian perspective there are no inviolable values that we ought to respect no matter what happens, but rather all values are to be compared with other possibly competing interests. As seen, this holds true not only if we compare an ordinary person and a cognitively impaired infant, but also if we compare the weight involved in the killing of diverse ordinary adult persons. In sum, we cannot in principle exclude the possibility of a person having a preference that is sufficiently weighty to justify the killing of another person whose preferences for continuing to live are weaker. As a possible preliminary reply, Singer might say that this is true but unlikely to be the case. As we shall see below, I will contend that this is in fact not a rare event, and that whatever the frequency, we should worry about the logic behind it.

III.2. Frustrating preferences impossible to fulfil

The second point to discuss concerning the overall tenability of a satisfaction-based interpretation of the value-theory of preference utilitarianism concerns the problem of the moralization of frustration.

¹¹ Singer, *Practical Ethics*, p. 107.

By the moralization of frustration of preferences I mean that, if the value of satisfaction-based utilitarianism depends solely on the satisfaction or frustration of preferences, we have no criterion by which to discriminate between preferences actively frustrated by a wrongful action perpetrated either intentionally or unintentionally by an individual, and the frustration of preferences caused by chance, natural events, unintended circumstances, or by the mere fact that a certain preference cannot be satisfied at all. Consider the following case. There are three teenagers: *A*, *B*, *C*. All love music and passionately want to become famous pianists. *A* has a natural talent for playing the piano and is a hard-working person. *B* is equally talented but lazy. *C* is a very hard-working person but is not talented at all. Suppose they start studying piano together and have an equally intense desire to become pianists. Building on this, according to the pure axiology of preference satisfaction, we may say that at this point in their lives thwarting their preference is equally wrong for all. But this is implausible because only *A* would be capable of actually becoming a pianist, while *B* would fail because of her laziness, and *C* because of her incapacity. Consider the thwarting of *C*'s desire to become a pianist. Would we say that such a thwarting is morally relevant? Would we call this a thwarting problem?

At this point Singer may respond that this is a problem only in a moral ledger model of accounting for the satisfaction of preferences. The moral ledger model suggests 'that we think of the creation of an unsatisfied preference as putting a debit in a kind of moral ledger that is merely cancelled out by the satisfaction of the preference'.¹² But this model is to be rejected, Singer argues, because in the end we should evaluate any life negatively, since in any life there are at least some unsatisfied preferences. Singer's favoured model for accounting for preference satisfaction in a life overall is the 'voyage model'.¹³ In this model, the value of satisfying preferences depends at diverse stages of life on the 'various amounts of hope and desire', as well as on how much 'time and effort have been invested in order to reach particular goals or destinations'.¹⁴ On this model, the wrongness of one's frustration depends on the amount of hope, commitment, and effort that has been put into the pursuit of a certain goal. This model may explain the difference between the frustrations of *A*, *B* and *C*. If somebody frustrates *A*'s preference there seems to be a moral problem, for *B* should only blame herself for her failure to achieve the desired result; while in the case of *C* there does not seem to be a real moral problem, for it is rather a problem of misplaced expectations and self-deception.

¹² Singer, *Practical Ethics*, p. 129.

¹³ Singer, *Practical Ethics*, pp. 129-31.

¹⁴ Singer, *Practical Ethics*, p. 130.

But the fact that *C*'s frustration is less of a wrong can be explained only if we include a correction of mere preference-satisfaction as a source of value. Preferences should be discounted in virtue of the overall probability of being satisfied. Thwarting my preference to become the most famous person in the world should be hugely discounted by the fact that this event is nearly impossible.¹⁵

Singer may respond to similar charges by appealing to a more normatively laden account of preference formation:

So somehow the preference utilitarian must give preferences a weighting which is independent of present intensity. The restriction usually suggested, following Butler's classical account of prudence, refers to reflection 'in a cool hour', based on an accurate knowledge of the relevant facts. If we build enough into this notion of full information, careful thought, and so on, we shall end up identifying preferences and interests. We shall then have a form of utilitarianism which seeks to maximize that special subset of preferences we call our interests. Among our interests, we would certainly give priority to long-standing, overarching preferences which (irrespective of how intensely they may be felt at a particular time) must be presupposed to make sense of the whole range of a person's activities.¹⁶

Can these considerations defend Singer's account from the charge I have just levelled? Probably. But the cost of embracing a fully informed and rational desire-based account of preference is high. I will not rehearse the many critiques levelled at this model by de Lazari-Radek and Singer in their book. I just want to add the following. The further disadvantage of employing a rationality-constrained account of preferences is that it would rule out animals from the domain of those capable of forming worthwhile preferences. If the preferences that count are only those that we would form rationally under a condition of full information, it follows that many standard human preferences are to be excluded, as well as the preferences of non-human animals. This seems an unjustified exclusion of the point of view of animals. If the preferences of animals were also to pass this rationality test, we would no longer know whether such preferences were really the

¹⁵ 'The implication of the objection to killing as the prevention of future preference fulfilment is that the strength of any preference will be an objection to extinguishing it only if that preference would otherwise be fulfilled. Where a preference would otherwise be fulfilled, its strength can indicate the amount of future positive value that its extinction depletes. But the strength of a preference that will not be fulfilled indicates the negative value that its contravention would constitute. Thus, to extinguish a slight or a moderate preference that would otherwise be fulfilled depletes future positive value. To extinguish an intense preference that would otherwise be contravened does not deplete future positive value. On the contrary, it prevents future negative value' (S. Uniacke, 'A Critique of the Preference Utilitarian Objection to Killing People', *Australasian Journal of Philosophy* 80 (2002), pp. 209–17, at 213).

¹⁶ P. Singer, 'Life's Uncertain Voyage', *Metaphysics and Morality: Essays in Honour of J. J. C. Smart*, ed. P. Pettit, R. Sylvan and J. Norman (Oxford, 1987), pp. 154–72, at 168.

preferences of non-human animals or just a human projection onto animals. It is also unclear how we could conceptualize an animal's preference formation under full information and rationality. Hence, the rationality and full information requirement cannot hold for animals and seems at odds with the overall intention of Singer's approach, which aims to be impartial among species and inclusive of all sentient beings.

III.3. Replaceability of self-conscious beings?

The third problem with preference utilitarianism concerns whether it can rule out the replacement argument in the case of self-conscious beings. Can we replace a self-conscious being, thus frustrating her future-regarding preferences, with another self-conscious being whose preferences will be satisfied? Singer answers this question in the negative, because the efforts of the first person will be lost:

It seems plausible to hold that this value, the value that is lost by a person's death after years of effort and striving towards a goal, is not going to be replaced by bringing a new person into existence, for then, as with the hedonistic version, there will be two lots of strivings for only one lot of fulfillment. Even if we do hold that there is some positive value in bringing into the world a being whose preferences will be satisfied, we can plausibly deny that the positive value of bringing a new being into the world is ever going to make up for such a loss.¹⁷

However, this response seems unconvincing for two reasons. First, the disvalue of frustration depends, as we have seen, on whether the frustrated preference would have been fulfilled. If it would not have been, no serious moral wrong can be predicated upon the mere fact of preference frustration. Second, this response unwarrantedly assumes that the preferences of the two persons (the first person, who will die, and the 'substitute') are equal. Singer's argument becomes weaker if we compare the preferences of two persons with unequal overall preferences and interests in continuing to live. As seen, such unequal preferences are the natural implication of this view and are also embraced by Singer himself.

Singer claims that self-conscious animals are not replaceable because their lives are biographical. Here, however, Singer conflates the argument against replacement with the argument based on the weight of the frustration or the satisfaction of preferences. The fact that a person's preferences are non-replaceable means simply that they cannot be substituted with the equally weighty preferences of another person. But this does not mean that they cannot be overridden.¹⁸ To appreciate this point, let us unpack the argument underpinning the

¹⁷ Singer, 'Life's Uncertain Voyage', p. 69.

¹⁸ Uniacke, 'A Critique of the Preference Utilitarian Objection to Killing People', p. 216.

idea that self-conscious animals cannot be replaced. This argument consists of two components. The first is that self-conscious animals have complex preferences regarding the future. Such preferences are weightier than other simpler and non-future-oriented preferences. Call this the *weight feature*. The second component is that such preferences are inextricably dependent on the individual life of a being, and in this sense cannot be replaced because there no other individual's preference is the same as the one we are considering. Call this the *idiosyncratic feature* of preferences. These two features do two different jobs in the argument. The weight feature is supposed to justify the duty not to kill a self-conscious being. The idiosyncratic feature of preferences blocks the possibility of replacing such a preference in case there is a supposedly equivalent substitute preference, but it does not tell us that such a preference cannot be overridden by another person's weightier preference. In a utilitarian framework, the interest or preference of a being can in principle be overridden by another individual's stronger interest or preference. To think otherwise would be to violate the consequentialist and monist nature of utilitarianism. If Singer's argument blocks, strictly speaking, the replacement of a self-conscious being by a being of equivalent interests, it does not rule out the possibility of killing a self-conscious being for the sake of providing benefit to another self-conscious being that is greater than the harm done to the killed being in virtue of the difference in weight of the preferences of the two individuals. This need not be a very outlandish case. Consider, for instance, that the only way for (α) (see [section III.1](#)) to continue living and pursue her very valuable preferences would be to receive a new heart. Suppose the only available donor is (β), who is very healthy but whose plans for life, if she has any, are far less valuable than (α)'s. What prevents us from saying that even a forced transplant is not justified?

Singer has two ways of rejecting this critique. The first is by saying that the fundamental preferences involved in the continuation of a self-conscious being's life are more or less the same across different individuals, and this justifies the duty not to kill a self-conscious being in order to promote the interests of another self-conscious being. But this is simply contradicted by the scalar nature of preferences and by Singer's explicit commitment to the idea that such preferences and interests vary greatly. The second possibility is that of holding that the preferences related to the continuation of one's life are irreducible and cannot be overridden by other preferences. But this runs counter to one of utilitarianism's basic tenets, namely the idea that there are no irreducibly different values and what counts is in principle (and in practice) measurable and comparable to other things with more or less value.

At this point Singer has two further options. Either he can suppose that, as far as matters of life and death are concerned, there is a threshold beyond which utilitarian calculation should not be employed – but this would end up being an abandonment of his approach – or he can admit that it is in principle justifiable to kill a self-conscious being for the sake of satisfying another self-conscious being's more weighty preference, but in practice this cannot occur.¹⁹

III.4. Moral ignorance and esoteric morality

We will see below that this disjunction between what is theoretically required and what is in practice justified is in its turn troublesome. For the moment, let us assess another argument that Singer has for preventing the unpalatable implications stemming from the satisfaction-based interpretation of his view. This is the 'don't know, don't kill' (DKDK) argument recently put forward by Guerrero:²⁰ 'If someone knows that she doesn't know whether a living organism has significant moral status or not, it is morally blameworthy for her to kill that organism or to have it killed, unless she believes that there is something of substantial moral significance compelling her to do so.' At first glance, this seems applicable to our case because if we don't know whether the overall preferences for continuing to live of an individual whom we know is a self-conscious being are sufficiently weighty to justify the prohibition of killing, we may still think that it is our duty not to run the risk of killing her. As explicitly argued by Guerrero himself, the DKDK principle is certainly compatible with different forms of utilitarianism. Singer *de facto* employs a prudential argument to extend the status of self-conscious animals to chickens without drawing its full implications.²¹ However, fully embracing this solution would have the disadvantage of making the distinction between conscious and self-conscious animals impalpable. After all, it seems plausible to think that the argument from moral ignorance extends well beyond the assessment of the weight of self-conscious individuals' preferences for continuing to live because it suggests we should probably not kill conscious animals either in cases where we might not in principle exclude the possibility of their being self-conscious.

¹⁹ It is worth remarking that my critique of Singer's argument against the killing of self-conscious beings considers only the strength of the preferences of possible individuals. It does not rely on the idea that Singer's act-utilitarianism is self-defeating or on the ideal case of a whole society following act-utilitarianism. On this, see P. Singer 'Is Act-Utilitarianism Self-Defeating?', *The Philosophical Review* 81 (1972), pp. 94–104.

²⁰ A. Guerrero, 'Don't Know, Don't Kill: Moral Ignorance, Culpability and Caution', *Philosophical Studies* 136 (2007), pp. 59–97, at 78–9.

²¹ Singer, *Practical Ethics*, p. 133.

Moreover, there are two further reasons why the DKDK principle would be unworkable on Singer's account. If we interpret the DKDK principle as attached to moral status, in virtue of the precautionary principle we should morally respect a number of beings about whose moral status we are ignorant. This restrictive understanding, however, leads us to a sort of deontological approach grounded on moral ignorance and prudence, which obviously cannot constitute a solution for Singer. If, on the other hand, we interpret the DKDK principle as an argument giving us at least a reason to factor into our moral consideration the mere possibility that a being might have a moral status, it would all depend on how we estimate and weigh such a possibility. It might be the case, indeed, that we estimate both a being's interest in living and the probability of its having a relevant moral status as very low, thus yielding a very weak reason against the killing of that being. If so, the DKDK principle would not add very much to our usual arguments or save Singer from the unwanted implications.

But the most powerful reason against the applicability of the DKDK principle is that it would force us to take our epistemic limitations very seriously, and in particularly more seriously than it would be possible to admit on Singer's act-utilitarianism. Indeed, if we concede this, the very possibility of having an esoteric morality (see also [section IV](#)), of factoring all individual preferences, and of having 'a point of view of the universe' would be radically jeopardized.

Along this line of thought, as a general rejoinder Singer can reply that the shocking effect of his arguments impacts on the critical level of morality, not the ordinary level of morality, which equally condemns all killing of people. On this point Singer is prepared to bite the bullet:

[I]f we take seriously the idea of life as a journey, the value of every life will vary according to the stage of the journey that the person has reached. So the wrongness of killing will be highly variable... At first glance it does seem wrong to vary the seriousness of a murder according to some judgement of the value of the life of the murdered person but viewing life as a journey does not compel us to do this. We can easily distinguish such practical matters as the criminal law from the philosophical views we hold about the value of life.²²

I will assess the overall plausibility of esoteric morality in [section V](#). Here it will suffice to say the following. The divergence between common-sense moral rules and esoteric morality is not per se a problem on Singer's account. However, for such a divergence to be acceptable there should be good grounds supporting the need for the common-sense rule. If there are not good grounds we might have reason to abandon such a rule (in our case that of considering all killing of

²² Singer, 'Life's Uncertain Voyage', p. 164.

diverse people an equally wrong act, despite the diversity of individuals' interests and preferences). My point here is that Singer does not provide good grounds justifying this divergence. He draws an analogy between the reason we have for considering all killing of people equally wrong under criminal law and the case for religious toleration. In both cases, we have pragmatic reasons that, however, cannot be justified at the philosophical and critical level:

Philosophically, it would be very difficult to defend the view that each religion had an equal claim to be regarded as true; but as a matter of practical politics, religious toleration is desirable We could take the same attitude in regard to the value of life: for certain purposes we could treat everyone's lives as being of equal value, even though we know that this is not really the case.²³

But the analogy with religious toleration is misplaced because religious toleration is not necessarily grounded in the idea that each religion 'has an equal claim to be regarded as true'. On the contrary, whatever the truth of religions, religious toleration can be justified on the basis of equal respect for persons and/or the equal liberty of anyone to profess her religion provided that it does not infringe others' rights. Modern states' religious toleration does not imply equal epistemic assessment of religions' claims to truth, but rather neutrality. Hence, religious toleration is not a second-best option to be pursued because we cannot impose the truth about religion; it is, rather, a principled choice.

A similar point can be made for the value of people's lives. Although all people have varying capacities – rationality, moral agency, sentience²⁴ – that determine the worth of their lives and their moral status, we may give equal consideration to all those who possess at least a certain relevant amount of such capacities.²⁵ This solution, based on the possession of a range of properties, can reconcile egalitarian principles with the undeniable fact that morally relevant properties are variable. Whether this solution is all things considered convincing is an open question and cannot be discussed here. However, this means that there is at least a principled and not merely pragmatic reason to uphold the equal moral status of persons, despite the variability of their morally relevant characteristics.

²³ Singer, 'Life's Uncertain Voyage', p. 164. On the same page, Singer says that in fact we already have such a critical attitude because 'we do not generally think the same efforts should be put into keeping seriously ill people alive irrespective of their age'.

²⁴ R. Arneson, 'What, if Anything, Renders All Humans Morally Equal?', *Singer and his Critics*, ed. D. Jamieson (Oxford, 1999), pp. 103–28.

²⁵ J. Rawls, *A Theory of Justice* (Cambridge, MA, 1971); I. Carter, 'Respect and the Basis of Equality', *Ethics* 121 (2011), pp. 538–71.

IV. OBJECTIVE VALUE THEORY: HEDONISM

To solve some of these problems we may abandon preference utilitarianism and think that the source of value lies in something that is independent of preferences. This is the route pursued by Singer in his latest book, co-authored with Katarzyna de Lazari-Radek. However, in moving from preference utilitarianism to Singer's more recent view, it is worth briefly discussing the argument put forward by Shelly Kagan. Kagan claims that in fact Singer was wrong 'in thinking that preference utilitarianism straightforwardly implies replaceability'.²⁶ He says that if we consider the preference satisfaction of two individuals, one (B) who replaces the other (A), and if we suppose that all their preferences are satisfied during their lifetime, the overall amount of preference satisfaction is less than in the case in which only the first individual lives her whole life without being replaced. This is so for the following reason. If we consider the time frame of (A)'s life, had she lived without being replaced, (B)'s preference satisfaction cannot make up for (A)'s lack of satisfaction of her preferences. However, this argument fails to support Kagan's claim convincingly, because he makes the very ad hoc move of considering only the life span of (A), who is replaced. Why should we restrict the scope of our concern only to this timeframe? If we also consider (B)'s lifespan, the overall preference satisfaction would be positive and the replacement justified. This restriction does not seem justified, because here we are working within the premise of the Total view, not the Prior existence view.

In general, Singer admittedly explains his change of view on the grounds that he has been persuaded that there are 'objective non-natural ethical truths' that define what is good for individuals independently of their preferences.²⁷ In this work, Singer and de Lazari-Radek subscribe to a hedonistic theory of what is good for conscious beings: 'Hedonism can be seen as an objective list theory with just one item – pleasure – on the list.'²⁸ To defend hedonism they point to the weaknesses and incoherence of the desire-based view. I will not go through these critiques. Rather, I will simply try to understand whether this latest view presents real advantages with respect to the issue at stake in this article. It is worth recalling that one of the advantages of the preference-based view was that it put forward an argument to avoid the unpalatable implication that the killing of a self-conscious individual could be justified on the grounds of the replaceability argument. Hence, we have to check whether hedonism

²⁶ Kagan, 'Singer on Killing Animals', p. 149.

²⁷ de Lazari-Radek and Singer, *The Point of View of the Universe*, p. 216.

²⁸ de Lazari-Radek and Singer, *The Point of View of the Universe*, p. 213.

does a better job of rebutting the idea that the killing of an individual might be justified.

Singer and de Lazari-Radek recognize that there may be situations in which it would be right to kill a person even against her will if we know, for instance, that she will experience a tremendous amount of suffering:

These conclusions are somewhat shocking, and so they should be. We need to have strong prohibitions against killing people against their will, because we are very unlikely to ever find ourselves in circumstances in which it is right to do that. If these acts of killing can be justified, they provide no basis for public policy, because the justification depends on them remaining secret, so that others do not become fearful that they will also be killed.²⁹

To defend this idea Singer and de Lazari-Radek put forward an overall endorsement of the esoteric morality first outlined by Sidgwick. Several critiques appealing to the moral principles of equality, autonomy and non-paternalism have been mounted against the idea of esoteric morality. In what follows, instead, I critically assess the plausibility of esoteric morality drawing on epistemic and applicability arguments. In particular, I show that the appeal to esoteric morality risks making the violation of the prohibition on killing more easily justifiable in practice than expected in theory.

V. ESOTERIC MORALITY AND ITS APPLICATION

Singer and de Lazari-Radek claim that for the practical purpose of common-sense morality we ought to comply with the absolute prohibition on killing, as well as other strict rules. I argue that in fact their argument aiming at enforcing the prohibition on killing rests on the unwarranted assumption that the justifiable killing of someone in secret is a very rare event. The problem with this argument is that it is questionable whether such events are so rare. To establish whether such events are really rare we need a criterion. But who is supposed to decide what this criterion is? Should we rely on the intuitive knowledge of some individuals of superior intelligence and calm disposition? This

²⁹ de Lazari-Radek and Singer, *The Point of View of the Universe*, p. 265. Here we may ask about the fate of merely sentient animals on this account. 'For the hedonist, the distinction between beings with self-awareness and those without it is not intrinsically significant. Perhaps it can be argued that self-aware beings are capable of greater pleasure than beings lacking in self-awareness, but presumably they are also capable of greater misery. . . . In explaining why killing humans is generally worse than killing animals (though not in all cases), the hedonist can once again appeal to indirect reasons: killing humans is likely to produce greater unhappiness among those close to the victim, and greater anxiety among others who fear being killed' (de Lazari-Radek and Singer, *The Point of View of the Universe*, p. 265). As is plain to see, on a hedonistic account replaceability is even more clearly endorsed.

perspective, indeed, relies on the possibility of a neutral and impartial individual who possesses the knowledge and a cool enough temper to examine hard cases. If one individual considered herself as such, and on this basis were to act accordingly, how could we assess her (in)capacity? And even more radically, we may ask how such a would-be expert could be sure of knowing the truth and being capable of accessing the critical level of morality. In what follows, I will provide some epistemic considerations to rebut the plausibility of this view.

First, suppose there are such experts. Here it seems we have a problem of establishing the expertise of a would-be expert by an audience of non-experts. To cut a long story short, in such a situation of novice–expert relations, Alvin Goldman very plausibly argues that the only reliable criterion for non-experts to ascertain the existence of experts and the quality of their epistemic skills is to check their track records, because the non-expert cannot appreciate and understand the epistemic value of the experts.³⁰ But de Lazari-Radek and Singer would object that I am assuming a publicity-based morality, which they reject.³¹ At the critical level of morality, there is just truth and no intersubjective recognition of morally and epistemically valid claims. But we have reasons to be suspicious about this position, because the principles of epistemic uncertainty applying to cases of ordinary morality are also likely to apply to the level of critical morality. Suppose there are people who know each other and consider themselves correct interpreters of this critical morality. Would they always agree? That seems implausible. Would they rely on an authoritative source to discriminate? That too is implausible because the only authority they would recognize is the authority of reason as a tool for reaching the truth. But, as we know from the epistemology of peer disagreement, it is very likely that equally competent persons who have the same evidence might still disagree on very important issues. Suppose that there are two competing groups of would-be enlightened utilitarians who seek to apply esoteric morality. Even if no ordinary person is supposed to check their expertise, how can each member of either group be sure that she belongs to the right group, correctly applying act-utilitarianism? Wouldn't these two putative utilitarian groups need a standard for assessing the mutual plausibility of their claims? And what would be the nature of such a standard? Wouldn't it require putting an end to strict secrecy?

³⁰ A. I. Goldman, 'Experts: Which Ones Should You Trust?', *Social Epistemology: Essential Readings*, ed. A. I. Goldman and D. Whitcomb (Oxford, 2011), pp. 109–33.

³¹ K. de Lazari-Radek, and P. Singer, 'Secrecy in Consequentialism: A Defence of Esoteric Morality', *Ratio* 23 (2010), pp. 34–58.

Second, suppose some objective moral knowledge exists, whether based on act-utilitarianism or otherwise. It is an obvious fact of the world that knowledge in morality, as well as in other domains, comes in degrees. People have diverse levels of knowledge in all domains depending on their varying intelligence, experience and commitment to the pursuit of truth. Building on this, the division of humanity between those who can access esoteric morality and those who cannot is like sharply dividing humanity between a technocratic elite and the mass of the uneducated. We might excuse Sidgwick for lingering on this idea given that it was commonplace in positivistic, as well as utilitarian, intellectual cliques in the nineteenth century. But we cannot equally excuse de Lazari-Radek and Singer for subscribing to such an implausible view of the distribution of knowledge and cognitive capacities in humankind.

Let us still assume that objective moral truth and knowledge exist. Consider a more plausible view in which moral knowledge is distributed on a continuum ranging from the restricted elite, which has full knowledge, and a large mass of unaware people. In between these two categories we might have at least the following types of persons who are ordered from the least to the most intelligent. Close to the mass there is Adam, who just knows that there is an esoteric morality because he once heard the elite in discussion. He understands that it is something different from ordinary morality but does not know what it amounts to. Then there is Eve, who came to know the general idea because she managed to read the Bible of esoteric morality. She grasps its general principles but cannot really understand its justification or how it works. Finally, there are two brothers, Abel and Cain, who not only read the Bible but also more or less understand it because they are more intelligent than Adam and Eve. However, they cannot apply it properly because they have never been invited to the meetings of the elite. Furthermore, Abel is good and Cain is evil. Suppose that, irrespective of the different levels of knowledge they possess, all four share the idea that esoteric morality entails the possibility of violating the rules of ordinary morality, whatever it means. How would they behave in light of this? Adam and Eve would probably be very confused and uncertain in applying the rules of ordinary morality. Abel and Cain, instead, might claim that their attempts to apply esoteric morality are justified. Given their different moral attitudes, Abel and Cain would probably err in different ways. But how can either of them claim that the other is wrong if they cannot rely on shared criteria to establish when esoteric morality is to be applied and what it demands? In all these cases, they would probably incur a number of mistakes and face moral confusion. Although we would need many more details to envisage this situation properly, and although depending on the

varying circumstances the overall utility might increase or decrease, we certainly face the problem of a theory that fails to reliably guide actions because ordinary people would not know what kind of rules (if any) should be followed. In general, it is unclear how this mismatch between epistemic and practical capacities would be more likely to bring about good outcomes. Hence, even if we cannot determine *ex ante* whether there would be an increase or decrease of overall utility, a failure to guide action is certainly an unconvincing feature of esoteric morality.

Certainly, de Lazari-Radek and Singer would rejoinder that a failure to apply a correct moral theory (act-utilitarianism) does not count as grounds for rejecting its validity. However, the assumption that only a few people with a highly developed epistemic level are legitimately supposed to apply esoteric morality is at odds with the obvious fact that people possess different cognitive capacities and different levels of understanding of this doctrine. The public nature of moral principles usually prevents this kind of problem because public criteria should be accessible to people who have diverse intellectual capacities.³² Such public criteria provide a stable ground for social interactions, the recognition of morality as a set of principles placing equal demands on all, and a way of making sense of mutual accountability. It is important to note that the condition of publicity I am assuming here does not need to be a condition of full publicity or publicity in the Rawlsian sense. Here I understand publicity as the condition that the content of moral principles should be epistemically accessible to average persons and that there should be some shared standards regarding how such principles should be applied. Although what I mean by publicity is included in the Rawlsian account of publicity, my condition is far less demanding. One may object that put this way my condition of publicity amounts to nothing more than a condition of ‘non-esotericity’ and, accordingly, it turns out to be too inclusive an idea. If so, fair enough – it is not a problem, because I do not want to make a general point on publicity and the condition is nevertheless sufficient for my claim to hold true. In sum, because of the epistemic uncertainty as to the level of critical morality and the scalar nature of knowledge, it is not clear how act-utilitarianism can function without some criteria of publicity, though they would run against the idea of esoteric morality.

Problems with the epistemic dimension of esoteric morality also have a bearing on the issue of feasibility. Singer and de Lazari-Radek

³² For this reason, the considerations I put forward are not liable to Eggleston’s critique against the necessity of the publicity condition. See B. Eggleston, ‘Rejecting the Publicity Condition: The Inevitability of Esoteric Morality’, *The Philosophical Quarterly* 63 (2013), pp. 29–57. On this see also B. Hooker, ‘Publicity in Morality: A Reply to Katarzyna de Lazari-Radek and Peter Singer’, *Ratio* 23 (2010), pp. 111–17.

place great emphasis on the importance of considering the feasibility and actual acceptability of utilitarianism. So, let us see how they incorporate this requirement into their theory. To make sure that a prescription following from the utilitarian principle can in fact produce better consequences, we have to ensure that all individuals to whom it applies, whether they can or cannot understand the utilitarian underpinning, would accept such a principle:

We entirely agree with Hooker that in proposing or promoting a moral rule for general acceptance in a society, it is vital to know whether it coheres with the prior moral convictions of most people, and hence has good prospects of easy acceptance, or clashes with these prior moral convictions, and so runs a high risk of rejection. In that sense, these moral convictions are data that we must take into account in deciding what we ought to do. But that is a very different thing from giving them probative force in deciding which normative theory we should accept.³³

So, ordinary morality must be plugged in to test the applicability of the utilitarian rule, but it cannot be employed to decide which moral principles are the valid ones. Besides the epistemic qualms we have just seen, the further problem with this role of esoteric morality lies in the fact that it is practically unstable. We have seen that there might be diverse people who could feel entitled to act according to esoteric morality without being capable. The lack of public knowledge of the conditions for the application of critical morality makes the situation structurally unstable. Either real cases in which critical morality may be applied against ordinary morality never occur, because the prudential principle prescribes a strict enforcement of the rules of ordinary morality in order to prevent misapplications of the ideas of critical morality – but if so, one may ask what the point of having a critical morality is, if after all it can never be put into practice. Or the application of critical morality, as envisaged above, is more widespread than expected by Singer and de Lazari-Radek, because, in the lack of public knowledge of the conditions for applying critical morality, a number of people who do not meet the epistemic and moral requirements to apply it do in fact consider themselves entitled to deviate from ordinary morality in a number of ordinary cases. But if so, the mass application of critical morality risks being detrimental to the pursuit of utilitarian goals because of the lack of proper capacities to apply utilitarian rules.

Singer's and de Lazari-Radek's reply to my point could be that the choice of options is misleading because, in fact, the correct response lies in between the frequent application of critical morality

³³ de Lazari-Radek and Singer, *The Point of View of the Universe*, pp. 305-6.

and the impossibility of applying it. But this intermediate position rests on the possibility that individuals capable of correctly applying act-utilitarianism actually exist and are successful in keeping act-utilitarianism secret. We have reasons to doubt such that figures actually exist and can manage this. First, that people vary greatly in their intellectual and moral faculties is certainly true. But this does not demonstrate that there are people capable of accessing and applying act-utilitarianism, and in particular its esoteric morality. Second, even if such people exist, it is not guaranteed that they will manage to keep esoteric morality secret and share its content only with their peers. This is so because keeping something secret is impossible if the group is not very restricted and because, even if the elite is capable of applying act-utilitarianism, it is not granted that they will also possess the capacity of individuating and appropriately selecting other members of the elite. After all, even Plato thought that the ideal city he outlined in the *Republic* was doomed because the eugenic system on which the selection of the elite is based is likely to be misapplied. One may say that this is not necessarily the case, because esoteric morality does not need an elite in charge of putting its content into practice. It might only need unrelated individuals who deviate from ordinary morality in order to put act-utilitarian critical morality into practice. Although this might in part be true, there are a number of issues concerning the ruling and coordination of a complex society that cannot be brought about by individuals alone. Hence, we must assume that for esoteric morality to be applied to collective issues there must be an elite of people who share the true content of esoteric morality and jointly act to bring it about when it is the case that the best utilitarian response would deviate from ordinary morality.

The implication is that we should put in question the applicability of act-utilitarianism. Hence, the esoteric morality based on a perfect judge does not seem to be available in practice. And, given Singer and de Lazari-Radek's commitment to the applicability of utilitarianism, this is a problem. It is particularly troublesome because there would be more people than expected who feel entitled to deviate from such a basic moral rule as the prohibition on killing. In sum, because of its epistemic problems and its applicability difficulties, act-utilitarianism's esoteric morality does not convincingly show that it can defend and enforce an overall prohibition on killing innocent people.

VI. CONCLUSION

To conclude, my argument in no way proposes that Singer's overall theory is untenable. It does not even prove that Singer's theory should be rejected because it conflicts with our moral intuitions. Rather, I

have cast some doubts on Singer's capacity to justify the usual moral wrongness of killing people. Such doubts concern both the preference-based account and hedonism. First, this is a problem because the preference-based account was also adopted to prevent the possibility of utilitarianism justifying the killing of people. Hence, this conclusion is troublesome not because it stands against ordinary morality, but because of its failure to live up to its commitment with respect to the distinction between conscious and self-conscious animals, and the prohibition on replacing the latter beings.

Second, the hedonistic account does not provide principled grounds for rejecting the justified killing of people but simply claims that it is a rare event in practice. But, as I have argued, because of Singer's reliance on Sidgwick's esoteric morality we don't know whether such cases are actually so rare. In sum, Singer's underpinnings of the prohibition of killing people face both theoretical and practical problems.³⁴

Federico.Zuolo@wis0.uni-hamburg.de

³⁴ The research for this article has been supported by the Italian Ministry of University and Research – FIRB 2010 – 'Feeding Respect: Food Policies and Minority Claims in Multicultural Societies', and by the Alexander von Humboldt Senior Research Fellowship 'Politics and Animals: Addressing the Disagreement about the Treatment of Animals'. I am grateful to Francesco Ferraro and Peter Niesen for their helpful comments on previous versions of this article.