ethic@

https://doi.org/10.5007/1677-2954.2025.e103162

15 YEARS ON: SAM HARRIS'S MORAL LANDSCAPE, A RETROSPECTIVE CRITIQUE

RAGNAR VAN DER MERWE[1]
(University of Johannesburg/SOUTH AFRICA)

ABSTRACT

In his best-selling book The Moral Landscape, Sam Harris develops a model for a universal ethics. His version of consequentialism has had some impact on the public, specifically on so-called new atheists. On the book's 15th anniversary, I look back at and retrospectively critique Harris's "science of morality" to see if it has stood the test of time. I first summarise the view, then identify various problems with it. Thereafter, I turn to arguably the most famous utilitarian, Peter Singer, for potential solutions to these problems. Since the publication of Harris's book, Singer and Katarzyna de Lazari-Radek have written The Point of View of the Universe. This book looks to Henry Sidgwick's The Methods of Ethics for answers to many of utilitarianism's perennial challenges, some of which Harris does not ostensibly overcome. My aim is to investigate whether Singer and de Lazari-Radek's utilitarianism (as informed by Sidgwick) can succeed where Harris's version seemingly does not.

Keywords: Sam Harris; Henry Sidgwick; Peter Singer; moral landscape; hedonistic utilitarianism.

[L]ooking at things ethically is a way of transcending our inward-looking concerns and identifying ourselves with the most objective point of view possible – with, as Sidgwick put it, 'the point of view of the universe'.

– Peter Singer (2011, 293).

Introduction

With several best-selling books and a popular podcast, Sam Harris has become a prominent voice in the public intellectual space. Although not a professional philosopher, he and his followers call him a "moral philosopher". Harris's ambitious book *The Moral Landscape: How Science Can Determine Human Values* (2010) is the bedrock for his unique brand of consequentialist ethics, one that has been especially impactful among the so-called new atheists. Harris expresses his "science of morality" as follows:

> I believe that there are right and wrong answers to moral questions in the same way that there are right and wrong answers to questions about biology. This commits me to what philosophers often call 'moral realism'. (2007, np)

Professional philosophers have largely dismissed Harris's ethics as amateurish and misguided. However, considering its popular impact, it might be fruitful to look back at *The Moral Landscape* on the 15th anniversary of its original publication. My goal in doing so is to investigate how Harris's view stands up to subsequent developments in academic ethics, specifically in utilitarianism. Despite his apparent lack of academic rigour, Harris does sometimes attempt to engage with the deeper philosophical arguments for and against the view he takes, and his argument is, therefore, worthy of critical analysis.[2]

As we shall see, the view expressed in *The Moral Landscape* faces significant challenges – challenges that Harris does not appear able to overcome. In particular, his central claim that science can *determine* human values is highly problematic; his arguments do not premise this bold conclusion. That said, some of the problems with Harris's view appear to be problems for utilitarianism broadly speaking, and some are, indeed, problems for any ethicist. After summarising his account, I turn to arguably the most famous utilitarian, Peter Singer, to see if he fares better in dealing with the pertinent problems.

Since Harris's book, Singer and Katarzyna de Lazari-Radek have written *The Point of View of the Universe* (2014),[3] which looks to Henry Sidgwick's *The Methods of Ethics* (1874/1962) for solutions to some of utilitarianism's perennial challenges.[4] These include (a) the problem of how to ground utilitarianism, (b) the problem of how to reconcile the conflict between self-interest and group-interest, and (c) the problem of how to define and measure utility. John Rawls, Robert Nozick, and Derek Parfit have also put forward famous challenges to utilitarianism – challenges that, as shall see, Harris struggles to deal with. My aim is to investigate whether utilitarianism – as developed after *The Moral Landscape* by Singer (sometimes with de Lazari-Radek and informed by Sidgwick) – can succeed where Harris seemingly did not. This is my novel contribution to the debate. I shall not simply dismiss Harris's ethical view in the way that most scholars have. Instead, I am going to investigate how it stands up to later developments in utilitarian ethics.

I conclude that some of the challenges are problematic for any moral theory aspiring to completeness, and therefore cannot be levelled against utilitarianism specifically. Singer offers plausible solutions to some of the

other challenges where Harris could not. Still other challenges are, on close analysis, not serious problems for contemporary utilitarianism. If so, then, although further philosophical hard work is needed, utilitarianism appears in relatively good shape moving forward.

This paper consists of four sections. In Section 1, I summarise Harris's ethics, paying special attention to the philosophical themes therein. In Section 2, I outline the major criticisms it has garnered from some professional philosophers. In Section 3, I look to Singer's recently developed version of utilitarianism to explore whether it has convincing answers to these criticisms. In Section 4, I summarise how and where Singer's view improves on Harris's view, and thereby on the status of contemporary utilitarianism in general.

1. The Moral Landscape

Harris generally underplays the importance of systematic, professional philosophy. Attacking traditional religious values and describing neuro-imaging studies on belief[5] take up much of *The Moral Landscape*. Nonetheless, while summarising his moral theory, I shall stay close to the philosophical implications and concomitant consequentialist/utilitarian themes.

1.1. HARRIS'S SUPPOSITIONS

Harris rejects moral scepticism, relativism, and religious ethics. He writes:

> [M]y (consequentialist) starting point is that all questions of value (right and wrong, good and evil, etc.) depend upon the possibility of experiencing such value. Without potential consequences at the level of experience – happiness, suffering, joy, despair, etc. – all talk of value is empty. Therefore, to say that an act is morally necessary, or evil, or blameless, is to make (tacit) claims about its consequences in the lives of conscious creatures (whether actual or potential). (Harris 2010, 62)[6]

In gist, this is a version of hedonistic utilitarianism (see Tännsjö 1998 for a book-length discussion and defence; see also de Lazari-Radek 2020).[7]

Harris's argument also rests on a naturalistic assumption about the mind. Conscious experience, he says, "entirely depends on events in the

world and on states of the human brain" (Harris 2010, 2). Therefore, "there *must* be a science of morality… because the well-being of conscious creatures depends upon how the universe is" (Harris 2010, 28 original emphasis; see also 2020 ch. 2). Scientific facts are then supposed to help us judge and improve the quality of sentient life through maximising well-being.

Harris also introduces a distinctive definition of 'science' as "our best effort to understand the universe" (2010, 29). 'Science' then appears to be synonymous with 'reasoned inquiry' (see also Harris 2016). This is an unusually broad definition. It seems that, for Harris, all genuine human knowledge counts as science (both rational *a priori* and empirical *a posteriori* inquiry constitute science). If so, then Harris's view begins to look like a science *and philosophy* of morality (rather than just a science of morality). We shall see in Section 2 what repercussions this has for Harris's overarching project.

1.2. A MODEL FOR MAXIMISING WELL-BEING

Having looked at Harris's starting suppositions, I now explicate the moral landscape model at the core of his thesis. He describes this landscape as a "hypothetical space" – a metaphysical model of

> real and potential outcomes whose peaks correspond to the heights of potential well-being and whose valleys represent the deepest possible suffering. Different ways of thinking and behaving… will translate into movements across the landscape. (Harris 2010, 7; see also 2020 ch. 2)

So, informed by the suppositions introduced in the previous section and our contingent socio-political climate, we survey a (metaphysical or imaginary) terrain containing highs of well-being and lows of suffering.[8] We then attempt to navigate toward the former and away from the latter. This model is a putative recipe for maximising the well-being of conscious creatures. According to Harris, one of the advantages of the landscape idea is that it can encompass multiple peaks of flourishing. Different individuals or societies can attain equal states of well-being by following alternative paths through the landscape (Harris 2010, 7–8).

The lowest valley on the landscape is what Harris called "the worst possible misery for everyone" (2010, 38–42; see also 2020 ch. 2). This is not a clearly defined notion with necessary and sufficient conditions. Instead, it is an intuitive hypothetical state that we can all agree should be avoided at all costs. One need only grant the following assumption: "The

worst possible misery for everyone is bad" (Harris in Dawkins and Harris 2011, np). Then, informed by our best science and accounting for cultural variety, we make decisions about how to proceed away from the worst possible misery and toward peaks of flourishing or maximum well-being.

1.3. Well-being, what is it?

Harris maintains that maximising well-being constitutes the only grounds for a workable ethics. He claims: "[A]ll efforts to describe morality in terms of duty, fairness, justice, or [virtue] draw upon some conception of well-being in the end" (Harris 2010, 33). Citing John Stuart Mill, Harris argues that Kant's *categorical imperative*[9] only qualifies as

> a rational standard of morality given the assumption that it will be generally beneficial… One could argue, therefore, that what is serviceable in Kant's moral philosophy amounts to a covert form of consequentialism. (Harris 2010, 199 fn. 10; see also 2020 ch. 2)

With regards to a definition of 'well-being', Harris claims that the diversity of peaks on the moral landscape allows that the meaning of 'well-being' is open to revision and (re)discovery. Nonetheless, he believes that any sensible definition will include notions of fairness and compassion, but future technological advances might introduce unexpected possibilities for human fulfilment and flourishing. Put otherwise, "our conception of 'well-being evolves'" (Harris 2010, 37).

1.4. The is-ought problem

Harris also mentions Hume's *is-ought problem*: The normative cannot be derived from or equated with the descriptive. Many consider Hume's distinction to be an unbridgeable gap for Harris's style of reductive ethical naturalism (e.g. Brown 2014; cf. Singer 1973). Harris, in contrast, maintains that a clear divide between facts (is) and values (ought) is illusory (see Endnote 2 below). Since epistemic and doxastic states reduce to brain states,[10] and, since objective knowledge claims are always value-laden, there is no distinction between the descriptive and the normative: "Scientific 'is' statements rest on implicit 'oughts' all the way down" (Harris 2011a, np). Therefore, "a science of morality should only pose a problem for you if you think a science of morality must be absolutely self-justifying in a way that no branch of science can be" (Harris in Dawkins and Harris 2011, np).

This supposed bridging of the is-ought gap is synonymous with Harris's claim that "science [related to what is] can determine human values [related to what ought to be]". As we shall see, commentators (rightly) seem to agree that this is the weakest aspect of Harris's ethics.

1.5. Some thought experiments

Harris also takes time to discuss some famous arguments against his brand of ethics. Many consider the following thought experiments to be stumbling blocks for consequentialism and/or utilitarianism.

Rawls's original position

First is Rawls's (1971) "justice as fairness" alternative to consequentialism. In Rawls's famous *original position* thought experiment, we consider the fairness of surrounding societal structures while imagining that we do not know which social status we have. We might well end up at the bottom of the hierarchy if a lottery arranged social structures. The lesson is that we must consider (even prioritise) the worst off in society. We should decide from behind a "veil of ignorance" what kind of socio-political structures we want to construct around us. Rawls maintains that we would make impartial and rational judgements in such a situation. In doing so, we collectively adjust our judgments and beliefs until they are in "reflective equilibrium" (i.e. stable, non-conflicting, and normatively applicable).

Harris grants that Rawls's thought experiment is "undeniably brilliant", but remains sceptical about whether anyone would accept the original position if it somehow led to bad consequences. For Harris, "[f]airness is not merely an abstract principle – it is a felt experience" (2010, 70). Fairness, he says, must, therefore, relate to consciousness and ultimately well-being. If so, then Rawls's notion of justice as fairness only really matters within a larger consequentialist concern for well-being.

Parfit's repugnant conclusion

Harris also engages with Parfit's (1984) much-discussed *repugnant conclusion*. If the utilitarian's concern is maximising total well-being, then a world in which many people have lives barely worth living is better than one in which a few people have ecstatic lives.[11] This is intuitively undesirable.

Harris appreciates the dilemma but insists that Parfit's challenge only shows how "certain moral questions could be difficult or impossible to

answer in practice" (Harris 2010, 72). Harris believes that we might one day solve these sorts of puzzles if we are guided by empirical data and honest conversation.

Nozick's utility monsters

Nozick (1974) imagines a scenario where there are creatures who get enormously greater well-being by devouring us than we lose by being devoured. He then asks if, on utilitarian grounds, it would be ethical for us to be sacrificed to these *utility monsters*. Most people intuitively answer "No".

Harris bites the bullet here and answers "Yes". He does, however, claim that other-worldly thought experiments of this nature are largely irrelevant to those of us who are concerned with actual (i.e. terrestrial) moral concerns. He also states that Nozick failed to consider the broadly negative psychological effects such a bizarre scenario would have on society at large (Harris 2010, 210–211 fn. 50).

1.6. Personal versus collective well-being

Harris recognises that "[o]ne of the problems with consequentialism in practice is that we cannot always determine whether the effects of an action will be bad or good" (2010, 67). How can one weigh the needs of one's own child against the needs of distant children, for example? This recalls the well-known 'self-interest versus group-interest' problem in moral philosophy. Harris cites studies showing how people find cooperation rewarding (Rilling et al. 2002) and concludes that a selfish individual might simply be ignorant of the happiness altruism can bring. It is possible for us to live happily with a general regard for the greater good of all sentient beings.

In sum, Harris argues in *The Moral Landscape* that the conventional is-ought dichotomy is an illusion: "Values are a certain kind of fact… they are facts about the well-being of conscious creatures. [As such] values reduce to the well-being of conscious creatures" (Harris in Dawkins and Harris 2011, np). For Harris, attaining well-being involves seeking and moving toward peaks on a landscape of possible states of flourishing. And because the "experience of conscious creatures is going to depend in some way on the laws of nature, there are going to be right and wrong ways to move across this continuum" (Harris in Dawkins and Harris 2011, np). Harris concludes that there can be a science of morality (or, as his book's subtitle suggests, "science can determine human values").

Philosophically, Harris's account is, in gist, an attempt to show that a version of utilitarianism – one that entails ethical realism, a reductive form of naturalism, and a central concern for well-being – serves as the only sensible grounds for ethical decision-making. In application, this plays out on an imaginary (or metaphysical) landscape model. This landscape contains a deepest valley – the 'worst possible misery for everyone' – but also peaks of flourishing. Informed by utilitarian suppositions and the empirical output of our best science (while taking contextual circumstances into account), we should then be able to successfully navigate the landscape.[12]

Harris acknowledges that his account faces certain problems, specifically the tension between personal and collective well-being. He is, nonetheless, convinced that even complicated moral questions have right and wrong answers. Just admitting this, he says, constitutes moral progress (i.e. a move away from ethical relativism and toward realism) (Harris 2010, 191).

2. Reviews and Unanswered Problems

Having summarised Harris's ethics, I now discuss some of the critical commentary that followed *The Moral Landscape's* publication.[13] Although many academics have simply dismissed the book's thesis as a naive form of scientism, others have written thoughtful reviews.

2.1. Critiques and book reviews

Despite complementing (a) Harris's rejection of scepticism and moral relativism and (b) his engaging survey of modern neuroscience, Kwame Appiah (2010) has criticised most of the rest of the book. Appiah targets Harris's outwardly *un*scientific supposition that we should increase the well-being of conscious creatures and argues that this undermines the book's general all-we-need-is-science motif (see also Kaufman 2010; Bělohrad 2011; Earp 2016; Champagne 2023; Percival 2023). Appiah also objects that Harris does not adequately deal with Parfit's repugnant conclusion or Nozick's *experience machine* thought experiment.[14]

Simon Blackburn (2011) has similarly objected that knowing all the scientific facts will not help us "organise" those facts in a way that promotes moral progress. In other words, we cannot cross Hume's is-ought gap as easily as Harris suggests (Radim Bělohrad 2011 and Mark Warren 2023 have criticised Harris at some length on this point; see also Kaufman 2010; Gert 2012; Meacheam 2012; Dunlop 2014; Earp 2016; Boornazian and

Diller 2023). Science also cannot instruct how we should (a) define and measure well-being and (b) balance personal versus collective well-being (see also Blackford 2010; Woien 2023). Like Appiah, Blackburn also criticises Harris's failure to engage with Nozick's experience machine thought experiment (see also Gordon 2023; Woien 2023).

On my reading, Thomas Nagel's (2010) review has been the most engaging. He commends Harris's promotion of ethical realism and his rejection of relativism and scepticism (see also Gordon 2023). Nagel concludes, nonetheless, that "the book is too crude to be of interest as a contribution to moral theory" (2010, np; see also Dunlop 2014). Specifically, he questions whether one can define 'science' as broadly as Harris does (Section 1.1) (see also Boornazian and Diller 2023; Champagne 2023; Percival 2023). Like Appiah, Nagel also points out the (non-scientific) utilitarian assumption at the foundation of an ethics that is supposed to be scientifically determined (see also Gert 2012). Nagel is also critical of Harris's quick dismissal of Rawls. He does not think that Harris has understood Rawls's theory of justice or competently dealt with the problem of personal versus collective good (see also Gordon 2023).[15]

2.2. Unanswered problems

Harris's landscape model has some appealing features. Firstly, the accommodation of *multiple peaks* of well-being eschews a potentially restrictive one-size-fits-all formula for well-being. Secondly, the notion of *the worst possible misery for everyone* provides a practical (albeit negative) reference point from which to navigate toward positive states of flourishing (if, of course, one accepts a form of consequentialism centred around well-being in the first place). Thirdly, Harris's core metaphysical landscape model appears systematic, potentially workable, and, in principle, quantitative. With suitable refinement, it could, perhaps, serve as a scaffold for some robust, applied ethical model (see also Meacham 2012).

That said, the above-discussed commentators have made some pointed criticisms.[16] Taking these into account, I identify the following seven unsolved problems (or unanswered questions) for Harris's version of utilitarianism/consequentialism:

P1. How do we ground utilitarianism?
P2. How do we reconcile the conflict between self-interest and group interest?
P3. How do we define and measure utility?
P4. How do we respond to Rawls's reflective equilibrium alternative?

P5. How do we resolve Parfit's repugnant conclusion paradox?
P6. How do we deal with Nozick's experience machine thought experiment (as opposed to the one involving utility monsters)?
P7. How exactly does science determine human values (or how exactly do we cross Hume's is-ought gap)?

The easiest to deal with appears to be P7. Almost every commentator has pointed out that Harris has failed to adequately answer this question (Section 2.1). There are, at least, three reasons to agree with the critics:

1. As mentioned (Section 1.1), Harris defines 'science' so broadly that it appears to be synonymous with 'knowledge' (both scientific and non-scientific). If so, then to say that "science determines human values" is merely to make the trivial claim that "knowledge determines human values".
2. Harris maintains that values inform science and that science can, therefore, inform values (Section 1). This is surely the case. The problem is that Harris mistakenly thinks that blurring the boundary between facts and values in this way means that science *determines* human values. However, facts and values being entwined or mutually affective does not entail the *one-directional* relationship of determination.
3. Harris's central premise that we should increase the well-being of conscious creatures is not *itself* determined by scientific output. Instead, this premise is derived from common sense. Science can (strongly) inform how we work out the pragmatics of valuing and increasing well-being, but it cannot "determine" the matter (as the subtitle of Harris's book suggests).

We can conclude that all Harris has shown regarding P7 is the truism that science can and should (strongly) *inform* (rather than determine) human values. If so, then we might be able to restate Harris's thesis in suitably subdued tones as follows:

> Assuming that we wish to maximise well-being and that well-being equates to quantitative brain states, some future neuroscience could strongly inform reasoned decisions about moral beliefs, values, actions, and the like.

This modified, conditional claim seems quite reasonable.

In any event, P1 to P6 might have solutions that Harris has overlooked, even if they present significant challenges. As mentioned in the introduction, I intend to look at developments in utilitarian philosophy after the publication of Harris's book to judge whether other utilitarians – specifically Peter Singer – have pertinent solutions.[17] The most famous philosophers in the utilitarian tradition are arguably Bentham, Mill, Sidgwick and now Singer (recently with Katarzyna de Lazari-Radek). Each has slightly different interpretations and versions of the general utilitarian approach. I shall, however, not present a history of utilitarianism here. I am specifically concerned with the six challenges at hand (P1–P6).

Written since *The Moral Landscape*, de Lazari-Radek and Singer's book *The Point of View of the Universe* (2014) deals with many of the topics I have been discussing here (see also de Lazari-Radek and Singer 2012). It therefore seems an apt place to look for the kind of answers I am looking for. A careful reading of Sidgwick's writings – specifically *The Methods of Ethics* (1874/1962) – has convinced Singer to shift from ethical anti-realism to realism and from preference utilitarianism (as outlined in Hare 1981) to hedonistic utilitarianism (originally developed by Bentham and Mill). This incarnation of Singer's ethics is interestingly close to the one Harris outlines in *The Moral Landscape*. This is another reason to think that investigating Singer's recently developed Sidgwickian view might prove fruitful to the purpose at hand (I am also going to cite more recent publications by Singer and de Lazari-Radek that touch on the relevant issues).

3. The Point of View of the Universe

Like Harris, Singer rejects scepticism and moral relativism. However, unlike Harris, he does not think that we can directly get an ought from an is (or that science can determine human values). In agreement with most of the commentators cited above, Singer maintains that we draw on normative reasoning rather than descriptive or empirical facts to determine ethical conclusions (even if the latter can and should inform the former). Both Harris and Singer also believe that moral claims can be *objectively true* and that moral rules and practices must ultimately be grounded in *consequences*. In discussion with Harris, Singer agrees that "any rule that reliably produced… misery would be dropped, and a different rule would be substituted" (Singer in Harris 2016). Singer also agrees that "intrinsic value is to be found only in conscious experiences" (albeit "not in all conscious experiences but only in positive ones") (Singer 2015, 147).

However, Singer considers *pleasure* (or sometimes happiness) to be the base unit of utility, and therefore has a slightly different emphasis

compared to Harris, who emphasises well-being (I discuss Singer's understanding of what pleasure entails in Section 3.3).[18] Nonetheless, the challenges P1 to P6 are relevant for either approach; both are, in gist, versions of hedonistic utilitarianism. Therefore, if Singer has solutions, then, by proxy, Harris might also. Similar to Harris, de Lazari-Radek and Singer set out in *The Point of View of the Universe* to, firstly, "argue that ethical judgments can be true or false", and, secondly, "to defend classical hedonistic utilitarianism" (2014, xiii; see also de Lazari-Radek 2020).

Like Harris and Singer, Sidgwick argues that nothing except *states of consciousness* has final value (de Lazari-Radek and Singer 2012, 238). Sidgwick is not concerned with "abstract philosophy"; ethics must involve "practical reason". He calls his version of utilitarianism "universalistic hedonism" or a "morality of common sense" (Sidgwick 1874/1962). According to de Lazari-Radek and Singer, Sidgwick's common sense morality implies that "the set of moral rules that we intuitively assume to be true [tends] to produce actions that maximize utility" (2012, 15). Moreover, an objective view – "the point of view of the universe" – is the "perspective of a rational being", and we would, therefore, do well to adopt it as we aim to maximise *the good* (de Lazari-Radek and Singer 2014, 378; see also Singer 2023, xi–xvii).[19]

With this in mind, let us now explore whether Singer (with the help of de Lazari-Radek and Sidgwick) might have solutions to P1 to P6. I shall outline Singer and co.'s proposed solutions in the rest of this section. In Section 4, I discuss those solutions' viability and how they match up to Harris's attempts.

3.1. P1: How to ground utilitarianism

According to de Lazari-Radek and Singer, we "seek clear and self-evident moral principles" that can lead us "cogently to trustworthy moral conclusions" (2014, 96). Regarding the problem of *first principles,* they explain how, for Sidgwick, one cannot establish foundational ethical presuppositions via strict deduction. To avoid abandoning morality altogether, we must give up attempts to rationalise such a foundation. Any workable self-evident rule can only achieve "the highest degree of certainty attainable" (rather than absolute certainty) (Sidgwick 1874/1962, 338; see also 1879).

De Lazari-Radek and Singer further argue that an ethic based on a non-consequentialist principle (e.g. Aristotelian virtue or Kantian duty) is logically circular. If one considers virtue, for example, to be the ultimate good, then one needs to know which qualities are, in fact, virtues. But, to

establish this, we have to know what we ought to do to gain these virtues. In other words, we need to appeal to the consequences of our actions. But, for that, we need an account of what is the ultimate good. If so, then defining 'ultimate good' as 'virtue' leads us around in a circle (de Lazari-Radek and Singer 2014, 234).

To avoid this circularity, de Lazari-Radek and Singer appeal to three moral axioms that Sidgwick proposes. These are supposed to be *self-evident* truths discerned by reflective reasoning (de Lazari-Radek and Singer 2014, 115–121):

A1: The principle of justice.
A2: The principle of rational self-love.
A3: (Most importantly) the principle of rational benevolence.

Similar to the Golden Rule, A1 appeals to our intuitive sense of fairness: No person has unmotivated special privilege. In other words, "whatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances" (Sidgwick 1874/1962, 379). However, this principle does not compel me to care about others or the greater good. Although it is rational to pursue my own interests, Sidgwick maintains that I should aim at my own good "on the whole". Universal good is the sum of individual goods, and this introduces A2: "[A]s a rational being I am bound to aim at good generally… not merely at a particular part of it" (i.e. not merely at myself) (Sidgwick 1874/1962, 382).

Via deductive inference, A1 and A2 conjointly lead to Sidgwick's primary axiom A3. Here,

> each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him. (Sidgwick 1874/1962, 382)

Put otherwise, "we are 'morally bound' to have as much concern for the good of any other individual as we have for our own good" except when (a) the other individual is objectively less deserving or (b) there is a limit to the requisite knowledge or ability to fulfil a requisite ethical act (de Lazari-Radek and Singer 2014, 133). This establishes what Sidgwick calls a "dictate of reason", which he believes should (*ceteris paribus*) motivate rational beings to perform suitably moral actions. In conjunction, Sidgwick's

three axioms are supposed to ground a utilitarian ethic. A1, A2, and A3 (with A3 as primary) are intended to offer a solution to P1.[20]

Despite this, de Lazari-Radek and Singer recognise that, in the end, all attempts to find *indisputable* first principles in systematic ethics will fail. Utilitarianism must be foundational on *rational* (but potentially disputable) intuitions about what is good. Singer thus maintains that to defeat relativism and scepticism, we must "attempt the ambitious task of separating those moral judgments that we owe to our evolutionary and cultural history, from those that have a rational basis" (2005, 251). If so, then solving P1 will involve an (ambitious) project of delineating rational intuitions from arational and irrational ones.[21]

3.2. P2: How to reconcile the conflict between self-interest and group interest

Can we reconcile individual egoism with universal altruism? Sidgwick (1874/1962) calls this self-interest versus group-interest dilemma "the dualism of practical reason", and considers it to be "the profoundest problem of ethics". Egoism is implicit in A2 (the principle of rational self-love), and altruism is implicit in A3 (the principle of rational benevolence). The problem is that self-love and beneficence tend to conflict (even when rationally informed).[22] We are forced to admit "an ultimate and fundamental contradiction in our apparent intuitions of what is reasonable in conduct" (Sidgwick 1974/1962, 508; see also de Lazari-Radek 2018).

Like Harris, de Lazari-Radek and Singer (discussing Sidgwick) explore whether we can argue that a selfish individual is simply irrational; an egoist simply does not realise that caring for others increases personal well-being (de Lazari-Radek and Singer 2014, 154–162). However, like Sidgwick, they conclude that this cannot be universally true. It is impossible to definitively demonstrate a robust (or necessary) connection between A2 and A3.

De Lazari-Radek and Singer (2012) have argued that A3 is consistent with the modern synthesis in evolutionary biology (e.g. Dawkins's 1976 'selfish gene' view). Like Harris, Singer (2015) also cites studies (e.g. Dunn et al. 2008) purporting to show how kindness, rather than self-centeredness, leads individuals to happier personal lives. It is, then, in our own interests to behave altruistically. Nonetheless, Singer and de Lazari-Radek recognise that, although such studies are encouraging, in the end, self-interest and group interest often pull in opposite directions.

3.3. P3: How to define and measure utility

According to de Lazari-Radek and Singer, Sidgwick's account needs a definition of 'the good' "or else it is empty of content" (2012, 27). Like Harris, Sidgwick claims that it is irrational to argue for some ultimate good (e.g. duty or virtue) without considering its relationship to sentience. For Sidgwick, when we take conscious minds into account, it is only reasonable to seek duty or virtue if they lead to pleasure. Ultimate good, therefore, equates to "desirable consciousness" (Sidgwick 1874/1962, 397–404) or what we might call *desirable conscious experience* (see also de Lazari-Radek and Singer 2014, 207–211, 252–253). This is the "greatest happiness" (or pleasure), which entails the "greatest attainable surplus of pleasure over pain" (Sidgwick 1874/1962, 120; see also de Lazari-Radek 2020; Singer 2023, 227–230).

De Lazari-Radek and Singer (2014, 240–244) argue, however, that Sidgwick struggles to present a practical technique for measuring pleasure. They go on to cite studies by Daniel Kahneman (1999, 2011) suggesting that we try to measure "instant utility" (see also Singer 2023, 223–226). Instant utility, explains Kahneman, "is best understood as the strength of the disposition to continue or to interrupt… current experience" (1999, 4). Here, pleasure is measured in terms of the subjective testimonies of individuals reporting on whether they wish to continue or discontinue some experience they are having. Similar to Harris, De Lazari-Radek and Singer suggest that, combined with empirical output from our best neuroscience, these reports could form part of a suitable measurement strategy (see also de Lazari-Radek 2020). Subjective testimony about someone's current disposition is combined with empirical data from the brain's pleasure centres.

Although promising, de Lazari-Radek and Singer recognise that these methods are only in-principle possibilities rather than tried-and-tested techniques. There also appear to be cases where utilitarian trade-offs are decidedly difficult – if not impossible – to weigh against each other. Singer (2015, ch. 12) asks, for instance, whether it would be better to donate $100,000 to a charity dedicated to slowing climate change or to one committed to restoring sight to one thousand blind people. It is not at all clear how we can weigh one against the other. Like Harris, Singer nonetheless insists that this does not mean that there is no objective answer. Instead, it "means that we don't have any way of knowing what the objective answer is because we do not have and cannot now get all the relevant facts" (Singer 2015, 145).

3.4. P4: How to counter Rawls's reflective equilibrium alternative

The view Rawls develops in *A Theory of Justice* (1971) is sometimes considered to be the most serious rival to utilitarianism. Harris, Sidgwick, de Lazari-Radek, and Singer maintain that moral knowledge must start from some grounded principle (or trinity of principles). Rawls's reflective equilibrium (Section 1.5), in contrast, represents a method by which we *construct* the best moral view (rather than *discover* a foundational moral axiom). For Rawls, our best moral knowledge is the most *coherent* possible set of beliefs we can construct.

In dealing with P4, de Lazari-Radek and Singer deny that there is, in fact, a clear distinction between moral foundationalism and Rawls's so-called moral anti-realism. That said, they reject coherentism in its pure form; a theory based on agreed intuitions cannot engender a sound ethics. Instead, it implies cultural relativism because different groups may come to different superficially coherent sets of moral beliefs. With no context-extrinsic rules, conflicting groups have no axiological recourse for settling moral disputes (de Lazari-Radek and Singer 2014, 100–104; see also Singer 2023, 8–11, 187–190; van der Merwe 2023, 2024, 2025a, 2025b).

Rawls sometimes distinguishes between narrow and wide reflective equilibrium (see also Daniels 1996). The former involves aligning particular moral judgments with moral principles. The latter goes further by also integrating relevant background theories (moral, social, metaphysical, epistemological, etc.). This means that it is not just about a mere coherence among intuitions and principles, but a coherence among all relevant considerations, including moral theory and empirical knowledge. This broader integration is supposed to guard against the relativistic collapse De Lazari-Radek and Singer are concerned about by bringing in external standards that can challenge and constrain parochial intuitions or local norms. In response, De Lazari-Radek and Singer question "whether wide reflective equilibrium does not then become so wide that there is no longer a contrast between it and any other way of justifying normative theories, including foundationalism" (2016, 112). In other words, either reflective equilibrium relies on contextualised (i.e. relativised) intuitions (narrow version), or it becomes indistinguishable from foundationalism (wide version), which is exactly what Rawls was trying to avoid.

In sum, an equilibrium of socially constructed guidelines only provides contextualised intersubjective agreements, not the objective moral truth needed to both ground an ethical system and properly guide our collective action. De Lazari-Radek and Singer conclude that Sidgwick's rationalist method offers better prospects than Rawls's constructivism does.

3.5. P5: How to solve Parfit's repugnant conclusion paradox

How should we practically distribute utility? Bentham famously stated, "[E]verybody to count for one, and nobody for more than one" (in Sidgwick 1874/1962, 417). Sidgwick agrees but adds that we should increase "happiness [or pleasure] on the whole" (1874/1962, 411–415). By default, sentience contains a surplus of pleasure over pain.[23] Therefore, "supposing the average happiness [or pleasure] enjoyed remains undiminished, Utilitarianism directs us to make the number enjoying it as great as possible" (Sidgwick 1874/1962, 415). However, this leads to Parfit's repugnant conclusion (P5) (Section 1.5).

Singer notes that P5 is not only a problem for utilitarianism. Any ethical theory making claims that affect future generations (as any good ethical theory should) must deal with this issue. Singer goes on to offer the "prior existence" version of utilitarianism as a potential solution. If beings exist or

> will exist independently of anything we choose to do, we ought to make their lives as good as possible; but we have no obligation to try to bring about the existence of people who, but for our actions, would not have existed at all. (Singer 2015, 172–173; see also de Lazari-Radek and Singer 2014, 366–376)

Regarding the distribution of utility, the prior existence view considers only (a) the happiness/pleasure of those who already exist prior to a utilitarian calculus and (b) those who will exist in the future, irrespective of that calculation. Prior existence utilitarians have no obligation to bring *new* happiness/pleasure into the world (see also Singer 2011). This outcome potentially blocks the repugnant conclusion.

3.6. P6: How to deal with Nozick's experience machine thought experiment

Nozick's experience machine thought experiment (introduced in Footnote 13) is often cited as *the* knock-down objection to hedonistic utilitarianism. In response, de Lazari-Radek and Singer challenge the thought experiment's intuition-based premise that people will not want to plug into the experience machine. Although many may find the idea repulsive, de Lazari-Radek and Singer suggest that this has mostly to do with the creepy science-fiction aspects of the scenario (such as having electrodes inserted into one's brain). They go on to cite research by Dan Weijers (2013) showing how, if the thought experiment is presented in less

viscerally intrusive ways, then people opt to plug in (de Lazari-Radek and Singer 2014, 254–259).[24] Lazari-Radek and Singer conclude that utilitarianism is not seriously threatened by Nozick's challenge. If plugging in would lead to a blissful world of harmony and prosperity (i.e. maximum overall pleasure), then it is perfectly reasonable to do so (even if such a world is illusory[25]).

4. P1–P7, Final Analysis

I have looked at the challenges P1 to P7 that Harris left unaccounted for fifteen years ago in *The Moral Landscape*. In this section, I briefly sum up what we can conclude from the above inspection of Harris's and Singer and co.'s ethics. Have the latter succeeded where the former outwardly failed?

P1

The problem of grounding utilitarianism appears to be a predicament for all moral theories aspiring to completeness. Any categorical first principle seems vulnerable to counterexamples. However, if we give up on a grounding criterion (or criteria), then scepticism and relativism threaten. We must somehow find a firm footing when no such footing can be established without an ostensibly unsupported stipulation.

Although Harris is (rather hastily) dismissive of P1 (Section 1.1.), Sidgwick, de Lazari-Radek, and Singer attempt to ground their ethics in three axioms arrived at by "reasoned" intuition: A1, A2, and A3 (the principles of justice, self-love, and benevolence, respectively) (Section 3.1.). This represents a middle road of sorts between Kantian realism and Rawlsian anti-realism. Perhaps, it is the best we can hope for when it comes to theoretical grounding. We can, then, conclude that Singer and co. have, if not solved, at least offered a best-we-can-hope-for solution to P1. In any event, what matters for current purposes is that P1 is not specifically a problem for Harris and utilitarians. Instead, it applies to any moral theory that attempts to block scepticism and relativism.

P2

Like P1, P2 is a challenge for any ethical realist (and arguably for all ethicists). How to reconcile self-interest with group interest – "the profoundest problem of ethics" – awaits a definitive resolution. As mentioned, Harris (Section 1.6) and de Lazari-Radek and Singer (Section

3.2) find studies showing how we feel satisfaction when helping others convincing in this regard. However, there is, of course, a limit to how much individuals will sacrifice for the good of the group. Regardless, it seems unfair to level P2 specifically at Harris or at utilitarians in general. As before, any robust ethical theory will have to grapple with this challenge.

P3

We have seen that Harris advocates for the pursuit of well-being. Sidgwick, de Lazari-Radek, and Singer roughly agree that moral rules must answer to the hedonistic utility of pleasure. They also seem to agree that an ethics that consistently made everyone miserable would not have any normative weight. In any event, P3 relates to the problem of how to define and measure the relevant utility (whether well-being or pleasure).

As discussed, Harris has been largely unconvincing in this regard (Section 1.4). Indeed, he seems reluctant to define 'well-being'. Sidgwick, de Lazari-Radek, and Singer roughly define 'pleasure' as 'desirable consciousness' (or 'desirable conscious experience') (Section 3.3). This seems plausible enough. De Lazari-Radek and Singer have also suggested that pleasure might, in principle, be measured through a combination of subjective reporting and applied neuroscience (Section 3.3). This gives their hedonistic utilitarianism some potential empirical stability. We can conclude that de Lazari-Radek and Singer have, at least, (a) given a plausible definition of their preferred base utility and (b) offered a plausible and potentially actionable method for measuring utility (even if the results remain to be seen).

P4

At first blush, Rawls's reflective equilibrium model appears to threaten utilitarianism's viability. Closer inspection has, however, suggested otherwise. Harris has been explicit that his project is specifically an attempt to provide a systematic account of foundational ethics. However, his realist starting assumption and associated quick dismissal of Rawls (Section 1.5) are disappointing for a thesis that purports to be comprehensive.

De Lazari-Radek and Singer have argued persuasively that reflective equilibrium can easily collapse into a kind of relativism (or radical contextualism) (Section 3.4).[26] There are different outwardly coherent sets of intersubjective beliefs. And, without any meta-criterion (or meta-criteria) for deciding between them, we might have to tolerate (or even encourage) conflicting moral attitudes.

Rawlsian contextualism further struggles to answer important universal questions like "What is *the* good (rather than many different kinds of good)?" and "How should humanity as a whole (rather than situated groups of humans) behave?" Arguably, reflective equilibrium thus falls outside the ethical project utilitarians are concerned with. It seems fair to conclude that Singer and co. (if not Harris) have adequately dealt with P4. Their sidelining of Rawls's approach appears justified.

P5

Parfit's repugnant conclusion presents a pragmatic challenge to utilitarians. As noted (Section 1.5), Harris is not sure how to deal with this issue. Singer, in contrast, has grappled at length with the paradox in his writings (e.g. 2011, 108–109; de Lazari-Radek and Singer 2014, 366–376). As mentioned (Section 3.5), P5 appears to present a predicament for any ethical theory contemplating the future of humanity. Like P1 and P2, P5 is, therefore, not a challenge specifically for utilitarianism. Regardless, Singer's *prior existence view* seems a viable response to the repugnant conclusion, even if it faces challenges of its own (see Singer 2011, 303–304; de Lazari-Radek and Singer 2014, 368–377). In sum, we can conclude that Singer and de Lazari-Radek have a possible (if incomplete) solution to P5.

P6

Although he discusses Nozick's utility monsters (Section 1.5), Harris (rather surprisingly) does not mention the experience machine thought experiment in his writings. As such, we cannot know for sure what his response would be. However, given his willingness to sacrifice himself to a utility monster (Section 1.5), we can assume that he would join Singer in plugging into the experience machine. Moreover, as noted (Section 3.6), if Nozick's thought experiment is suitably rephrased (without unpleasantries such as electrodes in the brain), then many people opt to plug in. It seems reasonable to embrace the opportunity to live in a dream of illusory bliss rather than endure the hard knocks of real life. If so, then it appears that P6 does not pose a serious threat to contemporary utilitarianism.

P7

As noted (Section 2.2), Harris appears to have failed when it comes to P7. He has not shown how science can determine human values (on any standard definition of 'determine'), and he has not shown how we can derive

an ought from an is (at least not in the straightforward manner he suggests). Instead, it appears that he has only affirmed the truism that science can and should (strongly) *inform* human values.

As Sidgwick notes, "there must be a gap in all such reasonings – where the notion 'ought' is introduced – which does not admit of being logically bridged over" (1879, 108). Singer (in Harris 2016) agrees that the is-ought gap stands despite Harris's ambitious, but ultimately misguided, attempt to bridge it.

Conclusion

I have looked back at the ethics Harris developed 15 years ago in his bold and popular, yet ultimately underdeveloped, book *The Moral Landscape*. With the help of some critical reviews by professional moral philosophers, I identified the problems P1 to P7. I then looked to de Lazari-Radek's and Singer's recent work (as informed by Sidgwick) for potential solutions.

In brief, I noted that contemporary utilitarianism has plausible (if tentative) answers to P1 (grounding utilitarianism), P3 (define and measure utility), and P5 (the repugnant conclusion). Further refinement is needed, but Singer and co.'s proposed solutions seem to present a solid platform to work from when it comes to these three challenges.

P2 (self-interest versus group interest) is a problem for all moral theories seeking completeness rather than being specific to utilitarianism. Perhaps personal versus collective well-being will forever remain in tension. In any ethically successful system of interacting persons, there might have to be a complex, constitutional balance between personal ambition (exemplified in, e.g., free speech and open markets) and group altruism (exemplified in, e.g., compulsory taxation and income redistribution).[27] I also noted that P4 (reflective equilibrium) and P6 (the experience machine) do not seem to pose serious challenges for the kind of utilitarianism Singer and de Lazari-Radek have developed since *The Moral Landscape*.

Regarding P7, it seems uncontroversial that Harris has failed to defend the bold claim that "science can determine human values" (or that we can derive an ought from an is). Singer agrees with this assessment and does not include any such claims in his theory (neither do Sidgwick and de Lazari-Radek).

In sum, contemporary utilitarianism (at least the Singerian version) appears to be in relatively good shape. Although further philosophical hard work is needed, it might represent a promising ethical model moving forward.

Notes

[1] Ragnar is a senior post-doctoral research fellow at The University of Johannesburg. His research interests are in the philosophy of science, philosophy of truth, pragmatism, and complexity science. Ragnar's PhD thesis focused on questions related to how we can have knowledge of the world given its complex nature and, most importantly, what this can tell us about scientific progress and scientific truth. Some of the journals where Ragnar has published include Axoimathes, Foundations of Science, and Journal for General Philosophy of Science. He has given talks at seminars hosted by, among others, Bath University, Kent University, University of Connecticut, and University of Vienna. Orcid: https://orcid.org/0000-0003-1038-758X

[2] Sandra Woien (2023) has recently compiled a collection of scholarly essays discussing different aspects of Harris's general philosophical view, including his ethics. This suggests that there might be some renewed academic interest in the topic. Harris himself has, however, not written much about the moral landscape or utilitarianism in recent years. His attention seems to have turned to existential threats related to AI and the supposedly impending singularity. As far as I can tell, he does, nonetheless, still subscribe to the general ethical view expressed in *The Moral Landscape*.

[3] Some of the ideas in that book are also contained in de Lazari-Radek and Singer's *Utilitarianism: A Very Short Introduction* (2017).

[4] David Phillips (2011) has also undertaken a notable book-length study of Sidgwick's ethics.

[5] Harris (2010 ch. 3) deals at length with neuroimaging studies that formed part of his academic research. He is especially interested in fMRI scans of peoples' brains when they form beliefs about the world. These studies show how the brain lights up in the same way whether forming beliefs about facts or values (e.g. Harris et al. 2008). For Harris, this suggests that there is a blurry boundary between facts and values.

[6] For Harris, "conscious creatures" or "sentient beings" include humans, so-called 'higher' animals (such as chimpanzees and dogs), and future AI systems (2014, 214–215 fn. 10; see also 2015).

[7] Harris recognises that his science of morality ultimately rests on an argument from first principles. He claims, nonetheless, that it could be falsified. His thesis would be disproved if (a) there is some alternative, cogent source of values that has nothing to do with the well-being of conscious creatures or (b) it is shown that well-being is unrelated to brain states (i.e. physical states that can be empirically investigated) (Harris 2010, 189–190).

[8] Harris (2020 ch. 2) sometimes refers to the moral landscape as a "metaphor".

[9] Kant's categorical imperative states as follows: "Act only according to that maxim whereby you can at the same time will that it should become a universal law".

[10] Harris does not consider *consciousness* (*viz.* sentience, awareness, or subjectivity) to be reducible in this way (2014, 216–217 fn 18; see also 2020 ch. 2).

[11] Alternatively, if one is concerned with *average*, rather than total well-being, then a world with one happy person is better than a world with billions of unhappy people. A utilitarian would then have to permit culling everyone except the happy person. This is why utilitarians tend to avoid talking about maximising average well-being (de Lazari-Radek and Singer 2014, 337–341; see also Sidgwick 1874/1962, 414–416).

[12] Quoting verbatim, Mark Warren (2023) summarises Harris's argument in the book as follows:

> 1.The well-being of conscious creatures is completely dependent on environmental and neurological states.
>
> 2.These states can be understood by science.
>
> 3.Moral truths are entirely reducible to truths about the well-being of conscious creatures.
>
> 4.Therefore, science is uniquely positioned to answer moral questions.

Warren considers Premise 3 to be the argument's Achilles heel because it is a philosophical supposition rather than an empirically derived scientific claim.

[13] James Diller and Andrew Nuzzolilli's (2012) review is the only one I can find that appears sympathetic to Harris's thesis in the book.

[14] Although Harris discusses Nozick's utility monsters, he does not mention Nozick's more well-known experience machine thought experiment. Nozick asks us to

> suppose there were an experience machine that would give you any experience you desired. Super-duper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting 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, pre-programming your life's desires? (1974, 43).

Most people intuitively find the plug-in option uninviting. Nozick concludes that we value more than maximizing well-being; we want to live in contact with reality. We

want to live rich, meaningful lives, even if doing so involves some inevitable hardship.

[15] Interestingly, most of the critics I have cited do not attack the metaphysics or utility of the landscape model itself. Marc Champagne (2023) has, however, noted that Harris presupposes that the highs and lows of the moral landscape have been differentiated before we begin navigating it.

[16] Following *The Moral Landscape's* publication and the resulting commentary, Harris (2011b) wrote an article titled "Response to critics". There, he mostly either repeats his arguments from the book or draws an analogy between well-being and *health*. He notes that, despite the difficulty in defining 'health', no one asks why maximising health is good in medicine. We can then, he says, presuppose the notion of maximizing well-being in a science of morality without a clear definition of 'well-being'. Ray Percival (2023) has persuasively criticised this argument by analogy.

[17] There are, of course, other problems for utilitarianism that are often cited in the literature. The impossibility of predicting the consequences of one's actions (Lenman 2000), the demandingness objection (Williams 1973; Hooker 2009), and the moral utility of non-human animals (Singer 1975; Kagan 2019) come to mind. However, the seven problems I have listed above appear most relevant to Harris's ethics.

[18] Harris prefers 'well-being' over 'pleasure' because he predicts that a future neuroscientist might be able to artificially vary our feelings of pleasure. We cannot, he says, have an arbitrarily variable state as our base unit of utility. He believes that well-being does not have this flaw because it is conceptually less precise (Harris 2010, 196 fn 20). See de Lazari-Radek (2018, 2020) for in-depth discussions about the challenge of defining 'well-being'.

[19] Interestingly all of Harris, Singer, de Lazari-Radek, and Sidgwick endorse the idea that suitably educated *moral experts* should dictate what society ought to do (Sidgwick 1874/1962, 373; Harris 2010, 36; de Lazari-Radek and Singer 2014, 21–21).

[20] Rawls (1974) argues that Sidgwick's professed ethical grounding represents an early version of his own reflective equilibrium schema. Sidgwick's foundational axioms 'hang together' as a putatively coherent and mutually supportive tripartite unit similar to the picture Rawls envisioned. Singer (1974) and de Lazari-Radek and Singer (2012) are, however, sceptical of Rawls's claim.

[21] There is ongoing debate about whether rationality in this context can be teased apart from the rest of our evolved (and therefore fallible) primate cognition. De Lazari-Radek and Singer (2012) say "Yes"; Street (2006), for example, says "No". See Vlerick and Broadbent (2015), van der Merwe (2022), and van der Merwe and Broadbent (2024) for reliabilist or piecemeal approaches that attempt to secure a middle way of sorts. Van der Merwe (2025a) argues that Gerd Gigerenzer's

ecological rationality might be capable of suitably constraining and informing moral decision-making and action.

[22] Parfit (2011) understands this dilemma to entail a choice between two questions: "What do I have most reason to do?" and "what ought I morally to do?" See also Barnhard Williams's (1973) famous critique of utilitarianism.

[23] David Benatar (2006) has put forward the counter-intuitive and much talked about view that bringing new life into the world increases overall suffering.

[24] De Lazari-Radek also suggests that

> our negative response to Nozick's imagined machine may be the result of a status quo bias, which can work the same in the opposite way – we would also be reluctant to change our current status if someone told us that we are already in an experience machine and we can be unplugged and return to reality (2020, 214; see also De Brigard 2010).

[25] In the Wachowskis' *The Matrix*, Cypher (played by Joe Pantoliano) opts to stay in the illusory world of the matrix where he can enjoy steak and fine wine rather than face the hardships of the real, outside world.

[26] Rawls implies as much when he states the following: "[O]bjectivity is not given by 'the point of view of the universe' to use Sidgwick's phrase. Objectivity is to be understood by reference to a suitably constructed social point of view" (1980, 570).

[27] Arguably, the world's most successful societies – seemingly in Scandinavia (Helliwell et al. 2020) – get this balance approximately right.

References

APPIAH, Kwame. Science knows best. *New York Times*, Oct. 1, 2010. Available at: https://www.nytimes.com/2010/10/03/books/review/Appiah-t.html. Accessed on: Apr. 4, 2024.

BĚLOHRAD, Radim. The is-ought problem, the open question argument, and the new science of morality. *Human Affairs,* v. 21, n. 3, p. 262–271, 2011.

BENATAR, David. *Better never to have been: the harm of coming into existence*. Oxford: Oxford University Press, 2006.

BLACKBURN, Simon. Morality without God. *Prospect Magazine*, 23 Mar. 2011. Available at:

https://www.prospectmagazine.co.uk/magazine/blackburn-ethics-without-god-secularism-religion-sam-harris. Accessed on: 24 Apr. 2024.

BLACKFORD, Russell. Review of the moral landscape, by Sam Harris. *Journal of Evolution and Technology*, v. 21 n. 2, p. 53–62, 2010.

BOORNAZIAN, Erik; DILLER, James. A moral compass that works. In: WOIEN, Sandra (Ed.). *Sam Harris: critical responses*. Chicago: Carus Books, 2023. Kindle Edition. Chapter 9.

BROWN, Campbell. Minding the is-ought gap. *Journal of Philosophical Logic*, v. 43, n. 1, p. 53–69, 2014.

CHAMPAGNE, Marc. My life gives the moral landscape its relief. In: WOIEN, Sandra (Ed.). *Sam Harris: critical responses*. Chicago: Carus Books, 2023. Kindle Edition. Chapter 2.

DANIELS, Norman. *Justice and justification*. Cambridge: Cambridge University Press, 1996.

DAWKINS, Richard. *The selfish gene*. Oxford: Oxford University Press, 1976.

DAWKINS, Richard; HARRIS, Sam. *Sam Harris and Richard Dawkins discuss: who says science has nothing to say about morality?* Richard Dawkins Foundation, 4 May 2011. Available at: https://www.richarddawkins.net/2011/05/who-says-science-has-nothing-to-say-about-morality-2/. Accessed on: 20 Aug. 2024.

DE BRIGARD, Felipe. If you like it, does it matter if it's real? *Philosophical Psychology*, v. 23, n. 1, p. 43–57, 2010.

DE LAZARI-RADEK, Katarzyna. On the notion of well-being. *Analiza i Egzystencja* v. 43, n. 1, p. 5–22, 2018.

DE LAZARI-RADEK, Katarzyna. What should a consequentialist promote? In: PORTMORE, Douglas W. (Ed.). *The Oxford handbook of consequentialism*. Oxford: Oxford University Press, 2020. p. 197-217.

DE LAZARI-RADEK, Katarzyna; SINGER, Peter. The objectivity of ethics and the unity of practical reason. *Ethics* v. 123, n.1, p. 9–31, 2012.

DE LAZARI-RADEK, Katarzyna; SINGER, Peter. *The point of view of the universe: Sidgwick and contemporary ethics*. Oxford: Oxford University Press, 2014.

DE LAZARI-RADEK, Katarzyna; SINGER, Peter. *Utilitarianism: a very short introduction*. Oxford: Oxford University Press, 2017.

DENNETT, Daniel. *Darwin's dangerous idea: evolution and the meanings of life*. New York: Simon and Schuster, 1995.

DILLER, James; NUZZOLILLI, Andrew. The science of values: the moral landscape by Sam Harris. *Behavior Analyst* v. 35, n. 2, p. 265–273, 2012.

DUNLOP, William. Review of the moral landscape: how science can determine human values. *Bioethical Inquiry* v. 11, n. 4, p. 557–561, 2014.

DUNN, Elizabeth W.; AKNIN, Lara B.; NORTON, Michael I. Spending money on others promotes happiness. *Science* v. 319, n. 5870, p. 1687–88, 2008.

GERT, Bernard. Neuroscience and morality. *Hastings Center Report* v. 42, n. 3, p. 22–28, 2012.

GORDON, David. Dark spots in Sam Harris's moral landscape. In: WOIEN, Sandra (Ed.). *Sam Harris: critical responses*. Chicago: Carus Books, 2023. Kindle Edition. Chapter 8.

HARE, Richard M. *Moral thinking: its levels, method, and point*. Oxford: Clarendon Press, 1981.

HARRIS, Sam. *The end of faith: religion, terror, and the future of reason*. New York: W.W. Norton and Co, 2005.

HARRIS, Sam. We are making moral progress. *Edge*, 2007. Available at: https://www.edge.org/response-detail/10416. Accessed on: 31 Mar. 2024.

HARRIS, Sam; SHETH, Sameer A.; COHEN, Mark S. Functional neuroimaging of belief, disbelief, and uncertainty. *Annals of Neurology* v. 63, n. 2, p. 141–147, 2008.

HARRIS, Sam. *The moral landscape: how science can determine human values*. London: Bantam Press, 2010.

HARRIS, Sam. Toward a science of morality. *HuffPost*, 6 July 2011a. Available at: https://www.huffpost.com/entry/a-science-of-morality_b_567185. Accessed on: 3 Mar. 2024.

HARRIS, Sam. Response to critics. *HuffPost*, 29 Jan. 2011b. Available at: https://www.huffpost.com/entry/a-response-to-critics_b_815742. Accessed on: 31 Mar. 2024.

HARRIS, Sam. *Waking up: a guide to spirituality without religion*. New York: Simon and Schuster, 2014.

HARRIS, Sam. What do you think about machines that think? Can we avoid a digital apocalypse? *Edge*, 2015. Available at: https://www.edge.org/response-detail/26177. Accessed on: 8 Aug. 2024.

HARRIS, Sam. What is moral progress? A conversation with Peter Singer. *Sam Harris*, 21 Oct. 2016. Podcast, MP3 audio. Available at: https://www.samharris.org/podcast/item/what-is-moral-progress. Accessed on: 3 Apr. 2024.

HARRIS, Sam. *Making sense: conversations on consciousness, morality, and the future of humanity*. New York: Ecco, 2020.

HELLIWELL, John F.; LAYARD, Richard; SACHS, Jeffrey D.; DE NEVE, Jan E. *World happiness report 2020*. New York: Sustainable Development Solutions Network, 2020.

HOOKER, Brad. The demandingness objection. In: CHAPPELL, Timothy (Ed.). *The problem of moral demandingness*. London: Palgrave Macmillan, 2009. p. 148-162.

KAHNEMAN, Daniel. Objective happiness. In: KAHNEMAN, Daniel; DIENER, Ed; SCHWARZ, Norbert (Eds.). *Well-being: the foundations of hedonic psychology*. New York: Russell Sage Foundation, 1999. p. 3-25.

KAHNEMAN, Daniel. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux, 2011.

KAGAN, Shelly. *How to count animals, more or less*. Oxford: Oxford University Press, 2019.

LENMAN, James. Consequentialism and cluelessness. *Philosophy and Public Affairs*, v. 29, n. 4, p. 342–370, 2000.

MEACHAM, Bill. Review of the moral landscape: how science can determine human values. *Philosophy Now,* v. 90, n. 1, p. 42–44, 2012.

NAGEL, Thomas. The facts fetish. *The New Republic*, 20. Oct. 2010. Available at: http://www.tnr.com/article/books-and-arts/magazine/78546/the-facts-fetish-morality-science. Accessed on: 30 Mar. 2024.

NOZICK, Robert. *Anarchy, state and utopia*. New York: Basic Books, 1974.

PARFIT, Derek. *Reasons and persons*. Oxford: Clarendon Press, 1984.

PARFIT, Derek. *On what matters: volume 1*. Oxford: Oxford University Press, 2011.

PERCIVAL, Ray S. The mantle of neuroscience. In: WOIEN, Sandra. (Ed.) *Sam Harris: critical responses*. Chicago: Carus Books, 2023. Kindle Edition. Chapter 6.

PHILLIPS, David. *Sidgwickian ethics*. Oxford: Oxford University Press, 2011.

RAWLS, John. *A theory of justice*. Cambridge, MA: Harvard University Press, 1971.

RAWLS, John. Kantian constructivism in moral theory: the Dewey lectures. *Journal of Philosophy*, v. 77, n. 9, p. 515–572, 1980.

RAWLS, John. Justice as fairness: political not metaphysical. *Philosophy and Public Affairs*, v. 14, n. 3, p. 223–251, 1985.

RILLING, James K.; GUTMAN, David A.; ZEH, Thorsten; PAGNONI, Giuseppe; BERNS, Gregory S.; KILTS, Clinton D. A neural basis for social cooperation. *Neuron*, v. 35, n. 2, p. 395–405, 2002.

SIDGWICK, Henry. *Methods of ethics*. London: Palgrave Macmillan, 1874/1962.

SIDGWICK, Henry. The establishment of ethical first principles. *Mind*, v. 4, n. 13, p. 106–111, 1879.

SINGER, Peter. The triviality of the debate over 'is-ought' and the definition of 'moral'. *American Philosophical Quarterly*, v. 10, n. 1, p. 51–56, 1973.

SINGER, Peter. Sidgwick and reflective equilibrium. *The Monist*, v. 58, n. 3, p. 490–517, 1974.

SINGER, Peter. *Animal liberation: a new ethics for our treatment of animals*. New York: New York Review, 1975.

SINGER, Peter. Ethics and intuitions. *Journal of Ethics*, v. 9, n. 3, p. 331–352, 2005.

SINGER, Peter. *Practical ethics*. 3rd ed. Cambridge: Cambridge University Press, 2011.

SINGER, Peter. *The most good you can do: how effective altruism is changing ideas about living ethically*. New Haven: Yale University Press, 2015.

SINGER, Peter. *Ethics in the real world*. Princeton: Princeton University Press, 2023.

STREET, Sharon. A Darwinian dilemma for realist theories of value. *Philosophical Studies*, v. 127, n. 1, p. 109–166, 2006.

TÄNNSJÖ, Torbjörn. *Hedonistic utilitarianism*. Edinburgh: Edinburgh University Press, 1998.

VAN DER MERWE, R. Rational decision-making in a complex world: towards an instrumental, yet embodied, account. *Logos and Episteme*, v. 13, n. 4, p. 381–404, 2022.

VAN DER MERWE, R. How pluralistic is pluralism really? A case study of Sandra Mitchell's integrative pluralism. *Theoria*, v. 38, n. 3, p. 319–338, 2023.

VAN DER MERWE, R. Stance pluralism, scientology and the problem of relativism. *Foundations of Science*, v. 29, n. 3, p. 625–644, 2024.

VAN DER MERWE, R. How Darwin can help post-structuralists maintain that apartheid was unconditionally unjust. *The Journal of Ethics*, v. 29, n. 2, p. 337–357, 2025a.

VAN DER MERWE, R. Perspectives and meta-perspectives: context versus hierarchy in the epistemology of complex systems. *European Journal for Philosophy of Science*, v. 15, n. 14, p. 1–20, 2025b.

VAN DER MERWE, R.; BROADBENT, A. A beginner's guide to crossing the road: towards an epistemology of successful action in complex systems. *Interdisciplinary Science Reviews*, v. 49, n. 5, p. 460–475, 2024.

VLERICK, Michael; BROADBENT, Alex. Evolution and epistemic justification. *Dialectica*, v. 69, n. 22, p. 185–203, 2015.

WARREN, Mark. A miserable argument. In: WOIEN, Sandra (Ed.). *Sam Harris: critical responses*. Chicago: Carus Books, 2023. Kindle Edition. Chapter 7.

WEIJERS, Dan. Nozick's experience machine is dead, long live the experience machine! *Philosophical Psychology*, v. 27, n. 4, p. 513–535, 2013.

WILLIAMS, Bernard. A critique of utilitarianism. In: SMART, James J. C.; WILLIAMS, Bernard (Eds.). *Utilitarianism for and against*. Cambridge: Cambridge University Press, 1973. p. 97-98.

WOIEN, Sandra. Another red pill. In: WOIEN, Sandra (Ed.). *Sam Harris: critical responses*. Chicago: Carus Books, 2023. Kindle Edition. Chapter 1.

Received/*Recebido*: 24/09/2024
Approved/*Aprovado*: 12/06/2025
Published/*Publicado*: 17/06/2025