Kumar, V. (2020), *The Oxford Handbook of Consequentialism* (OUP), Ed., Douglas Portmore

The Science of Effective Altruism

Victor Kumar

Boston University

Peter Singer might be the most influential moral philosopher since Plato. Among the general public, his work is immensely popular (at least when compared with other philosophical work in the analytic tradition). Almost single-handedly, Singer (1975) has sparked a revolution in attitudes toward non-human animals. Many people are vegetarians or vegans (or activists) principally because of him. Singer (2009, 2015a) has also made a powerful case for charitable giving toward people suffering from poverty and disease in the developing world. He's helped people grasp "the most good they can do."

Singer's career has followed in the tradition of other utilitarian social reformers like Jeremy Bentham and John Stuart Mill. These utilitarians responded to some of the most pressing social problems of their day. Bentham (1789) and Mill (1863) opposed racist and sexist discrimination. Their ideas are unremarkable nowadays, perhaps, but went decidedly against the grain of moral thought in 19th century Europe. Singer, by contrast, has focused on indifference toward non-human animals and the global poor. At the same time, Singer has also been undersensitive to the needs and interests of people with cognitive disabilities (see, e.g., Barnes 2016). His commitment to utilitarianism may have engendered certain moral errors (cf. Taylor 2017). Nonetheless, like Bentham and Mill, Singer has led the way on addressing important forms of social injustice, transcending society's existing limits.

Singer has not just influenced people who read his work or hear him speak. He has also sparked a growing and cohesive philosophy and social movement called "effective altruism." The movement is led by younger philosophers-cum-philanthropists like William MacAskill, Toby Ord, and others. These effective altruists follow in Singer's footsteps but aspire to leave an even bigger moral footprint.

Effective altruism is grounded in the idea that we should do the most good we can by using our resources in the most effective ways possible. If altruism is to be truly "effective," it must be guided by scientific evidence about how to make our money and our efforts go the furthest. Instead of picking what seems intuitively like the worthiest cause, one that imparts to the giver a "warm glow," effective

altruists argue that it is better to donate to international charities that will save the greatest number of lives per dollar. One controversy surrounding effective altruism, as we'll see, stems from the possibility that global development charities offer only drops in a bucket. The bigger problems, allegedly, are systemic. Charitable giving doesn't address these systemic problems, according to some critics, and may well have the unintended effect of sustaining unjust systems.

Effective altruism puts utilitarian ethical theory into practice. Its aim is to develop and enact *a science of doing good*, thereby integrating utilitarianism with science. However, Singer (2005) also claims that utilitarianism itself is scientifically justified. This is different from using science to apply utilitarian principles, as effective altruists intend. According to Singer, science supports utilitarian principles. This is a bold idea, to say the least, in part because of the well-known gap that divides "is" from "ought." Among philosophers, the idea that science supports utilitarianism is often met with extreme skepticism, if not scorn. Academics in other disciplines are often more receptive, however. Utilitarianism seems to enjoy broad appeal among scientists, or at least among those scientists who give thought to moral philosophy. This is a curious fact, indeed so curious that it begs for explanation.

My main aim in this chapter is to evaluate the scientific merits of effective altruism and utilitarianism. I'll construct a defense of effective altruism that identifies promise in its commitment to revise particular moral commitments on the basis of scientific evidence. I'll also argue, however, that effective altruism is more plausible when stripped of its supposedly utilitarian foundations. Science, alas, offers no support to utilitarianism itself. But the evidential role for science in effective altruism endows this philosophical position and social movement with significant moral and intellectual worth.

1. Utilitarianism and Effective Altruism

Utilitarianism is a beguiling ethical theory, as many ethics instructors can attest. One common argument for utilitarianism begins by asking what, if anything, has inherent value. Of all the many things people pursue, which are worthy in and of themselves, apart from their instrumental benefits? Why seek a large income? Why strive to be successful in your chosen profession? Why cultivate connections to friends and family? A tempting thought is that each of these pursuits is valuable insofar as it increases pleasure and alleviates suffering. Pleasure and the absence of suffering can be lumped together as "happiness." Thus, imagine that the connection between your pursuits and happiness is severed. If more money, greater professional success, or deeper relationships don't engender happiness, then it seems as though they are not worth pursuing.

Utilitarians suggest that simply by reflecting on your own experiences and activities, you can discover that happiness alone advances your interests. Other things advance your interests only indirectly—only by having an impact on pleasure or suffering. Once you see that happiness is so intimately connected to your own interests, it should then become clear that happiness matters just the

same when others experience it. Pleasure cannot become more valuable when you enjoy it than when I do, at least not from "the point of view of the universe" (Sidgwick 1907). So, to act morally you must consider how your action affects not just your own interests but also the interests of others. What matters is the impact on everyone's happiness, not the fact that other people are distinct individuals, nor that they live in a different place or time, nor even that they are members of a separate biological species.

Utilitarianism thus has two main tenets. One is a principle of inherent value: only happiness directly advances a person's interests. The other main tenet is a principle of equal consideration: everyone's interests matter equally, so long as they are capable of experiencing pleasure or suffering. Put these two principles together and you get utilitarianism—or "classical" utilitarianism. In every choice, to wit, the morally correct action is the one that has the highest expected value in terms of overall net happiness, while weighing the interests of every sentient being affected.

This is certainly a very brief presentation of utilitarianism and just one argument for it. We haven't considered other principles that are less salient but also essential to the theory. For example, classical utilitarianism is also committed to the principle of "maximizing" expected value rather than "satisficing" (cf. Lichtenberg, this volume). In light of this principle, morally correct actions should produce the most expected happiness, rather than simply more or "enough." We also haven't yet considered objections to utilitarianism. Some of the most pressing objections draw out implications of the theory for rights and partiality that are hard to swallow.

However, it isn't the purpose of this chapter to fully flesh out utilitarianism or offer a thorough evaluation of its merits. My central focus will be on effective altruism (hereafter "EA"). Over the course of this chapter I will show how utilitarianism leads to EA and explain what this view entails, in theory and in practice (this section). Next, I will defend EA from objections (sections 2 and 3). I will then argue that EA need not be grounded in utilitarianism (section 3). And finally, I will lay out why utilitarianism does not enjoy the empirical support that some proponents advertise (section 4). Utilitarianism entails EA, but EA does not entail utilitarianism—that is, as I'll argue, given a version of EA that most merits allegiance.

Singer's work is the impetus for EA, but younger scholars inspired by Singer's work have recently taken the reins. One of the most prominent effective altruists is William MacAskill. In *Doing Good Better*, MacAskill (2015: 11) unpacks EA as follows:

> Effective altruism is about asking, "How can I make the biggest difference that I can?" and using evidence and careful reasoning to try to find an answer. It takes a scientific approach to doing good. Just as science consists of the honest and impartial attempt to work out what's true, and a commitment to believe the truth whatever that turns out to be, effective altruism consists of the honest and impartial attempt to work out what's best

> for the world, and a commitment to do what's best, whatever that turns out to be.

Remember that EA is both a philosophical position and a social movement. As a philosophical position, EA says that we should take "a scientific approach to doing good." That is, if we have the time and if available evidence is not already decisive, we should rely on scientific evidence to determine how we can do the most good. (Whether EA says that this exhausts our moral obligations will be discussed later in section 3.) As a social movement, EA is a collection of individuals and organizations committed to following this principle. In this essay, "EA," unless qualified, denotes the philosophical position. However, I will sometimes distinguish the position from the movement, theory from practice. For example, I will argue that some criticisms of the practice do not undermine the theory; I will also suggest how the practice might be reformed so as to better live up to the theory.

Utilitarianism, it seems, straightforwardly entails EA. The entailment holds given the background assumption that, as seems correct, science offers the best evidence about the impact of one's actions on the interests of others. MacAskill (2015: 12) thus argues that an effective altruist should seek scientific answers to a number of particular questions. For example, who will be impacted by your actions and by how much? What are alternative courses of action and what would their effects be? What are the chances that your action will succeed in its aims? Scientific tools can address these questions, giving answers that are needed to best apply the utilitarian principles of inherent value and equal consideration.

Like many other effective altruists, MacAskill's main focus is on the science of doing good in the developing world. This focus isn't arbitrary. People in developed nations are members of the "global 1%" (2015: 15-25). Thus, in light of the diminishing marginal utility of wealth, global charity is much more effective than charity within developed nations. In terms of what some economists and medical ethicists call "quality-adjusted life years," (QALYs) "[t]he same amount of money can do [roughly] one hundred times as much to benefit the very poorest people in the world as it can to benefit typical citizens of the [developing world]" (22). Unless one happens to be a multi-millionaire, a charitable donation can only be relatively small—a mere drop in the bucket. However, "[i]t's the size of the drop that matters, not the size of the bucket, and, if we choose, we can create a [relatively] enormous drop" (25).

Having laid out the theoretical underpinnings of EA, MacAskill goes on to canvass relevant scientific evidence. For example, he outlines which international aid charities have an established record of cost-effective work (2015: 121-27). These include GiveDirectly, which offers cash transfers to poor households in Kenya and Uganda, no strings attached; Development Media International, which produces radio programs that educate families about basic health measures in several African countries; and the Against Malaria Foundation, which provides bed nets that protect people from mosquitoes and malaria in sub-Saharan Africa. All of these charities have empirically demonstrable efficacy and can scale-up their work on the basis of further donations.

MacAskill also argues that scientific evidence leads to a number of counter-intuitive conclusions. Just as a science of the material world overturns many widely held factual positions, a science of doing good overturns many widely held moral positions. For example, MacAskill criticizes a number of ways in which people engage in "ethical" consumption of consumer products. Some people in wealthy countries favor purchasing sugar, coffee, and other foods only through "fair trade" companies that provide employees with a decent wage. But since it is only companies in relatively well-off countries that can afford to adopt fair trade practices, MacAskill argues, the result is that workers in relatively disadvantaged countries lose employment. Better, then, to buy cheaper products that aren't fair trade and donate the difference in price to effective charities (2015: 132-5). MacAskill also criticizes certain "green" lifestyle choices, like turning off lights, refusing to use plastic bags, and buying locally sourced goods. Empirical evidence suggests that each choice is relatively ineffectual. Much more effective is carbon offsetting. Instead of consuming products that entail fewer greenhouse gas emissions, it is better to subsidize projects that reduce carbon emissions elsewhere (135-40).

MacAskill goes on to argue that EA provides a useful framework for choosing a career (MacAskill 2015: ch. 9). If you want to make a difference, the most effective career is not necessarily in the non-profit sector, even if the company you choose to work for happens to be very effective. If the person who would have been hired instead of you is nearly as skilled, then you aren't making much of a difference (155). MacAskill argues that one possible option is "earning to give" (163-4). If you have the right skills, it might be better to enter the financial sector instead of the world of non-profits, earn a much higher salary, and give away a large proportion of your income to charities that will do more good than you could have accomplished through your own non-profit work.

So far in this chapter, I've outlined one argument for utilitarianism, defined EA, and shown how utilitarianism entails EA as a philosophical position. I've also described EA in practice by recounting some of the recommendations that seem to issue from a science of doing good. In the rest of the chapter I'll turn to the business of critically evaluating EA. I'll begin by evaluating EA on its own terms before returning to its connection with utilitarianism. We'll have to wait until the final main section of the chapter before we consider scientific arguments for utilitarianism and evaluate their merits.

2. The Structural Objection

Casual assessments of EA are sometimes anchored in particular recommendations given by effective altruists, like those recommendations rehearsed at the end of the previous section. Thus, for instance, some people who have a positive opinion about EA may defend it with an example. Surely it's a good idea to donate to the Against Malaria Foundation instead of wasting money on Toys for Tots? However, an apparently sensible recommendation like this is not necessarily what comes to mind for critics. People who have a negative opinion about EA

sometimes appeal to a different example. Isn't it a bad idea to become a hedge-fund trader, instead of working directly with disadvantaged communities, simply because you would then have more surplus income available to donate to charities? Might this not be a betrayal of your values?

It is tempting to scrutinize particular recommendations given by effective altruists. Perhaps, for example, there is evidence that people who enter the financial sector are likely to forget about the noble ambitions that first led them there. A science of doing good must, of course, be open to this possibility and therefore willing to abandon this recommendation (cf. MacAskill 2015: 165-7). But this just shows that to assess EA as a philosophical theory we must go beyond examples of the practice. We can't let applications of the theory stand in for the theory as a whole.

One recurrent criticism of the practice of EA lies in its focus on advancing the interests of people in the developing world through financial aid. Many people are intensely skeptical about the value of international aid organizations. Some aid programs are ineffectual or worse. In addition, some programs are captured by the interests of corrupt institutions, whether in the countries where aid originates or in the countries where aid is targeted. However, MacAskill (2015: 47-53) claims that it's a mistake to focus on the worst cases. Overall, he argues, international aid has been an enormous success. It suffices that the best organizations have been extremely effective. MacAskill thus argues that the eradication of smallpox alone is worth all the money spent on aid programs (46). Arguments against donating to bad aid programs do not count against donating to good aid programs.

A deeper criticism of EA lies underneath the surface here—which I will call the "structural objection." The objection begins by arguing that suffering in the developing world stems from massive global inequalities in power. These inequalities rest on local and international institutions and the hierarchical social structure they engender. Effective altruists like MacAskill want to work with these institutions, instead of reforming them. According to the structural objection, however, the root of suffering in the developing world is systemic and structural inequality that gives wealthy nations outsized power and control in international affairs. This type of inequality cannot be opposed through financial aid, but only through political action, not individually but collectively. Drops, even enormous ones, will have limited value if the bucket is lopsided by design.

A number of critics have given voice to this objection. Amia Srinivassan (2015) expresses it lucidly in her gripping and wide-ranging review of MacAskill's book:

> Effective altruism, so far at least, has been a conservative movement, calling us back to where we already are: the world as it is, our institutions as they are. MacAskill does not address the deep sources of global misery—international trade and finance, debt, nationalism, imperialism, racial and gender-based subordination, war, environmental degradation, corruption, exploitation of labor—or the forces that ensure its reproduction. Effective altruism doesn't try to understand how power works, except to better align itself with it.

Srinivassan is more sympathetic to EA than some critics. She shares with effective altruists a similar perspective about the world's moral problems and accords them a similar priority. But she argues that the structural or systemic bases of the problems cannot be addressed through charity (see also, e.g., Herzog 2016).

Jeff McMahan (2016) does not find the structural objection quite so compelling. His response to the objection begins by pointing out that a moral agent does not have direct influence over social and political institutions. What she can directly control are her own actions and efforts. She may attempt to reform global economic institutions or she may take direct charitable action. It is best to do both (McMahan 2016):

> Yet there has to be a certain division of moral labor, with some people taking direct action to address the plight of the most impoverished people, while others devote their efforts to bringing about institutional changes through political action. To suppose that the only acceptable option is to work to reform global economic institutions and that it is self-indulgent to make incremental contributions to the amelioration of poverty through individual action is rather like condemning a doctor who treats the victims of a war for failing to devote his efforts instead to eliminating the root causes of war.

Let's linger on McMahan's last point. In "Famine, Affluence, and Morality," Singer (1972) famously asked readers to imagine the following case. You are walking by a shallow pond and see a toddler drowning. You can rush in, but doing so will ruin your very expensive new clothes. Are you obligated to save the child? Of course! Having taken the bait, however, you are now on the hook. In fact, you have the opportunity to save a child's life in the developing world for a financial cost that is similar (very roughly) to the price of your expensive clothes. The child is remote instead of nearby, starving instead of drowning, but the differences between the cases do not seem to be morally relevant. Every day that you choose material luxuries over charity you are effectively refusing to save a drowning child.

Provided that Singer's analogy is sound, proponents of the structural objection seem to be reasoning as follows: "You are walking by a shallow pond and see a child drowning. Are you obligated to save the child? No! Not if the child is drowning because of deeper structural causes, say, because the local government refuses to erect barriers around the pond. If that's the case, you should leave immediately and lobby the government to make structural changes, leaving the child to die in favor of more important problems." Effective altruists are sometimes held in suspicion for taking a heartless, mechanistic approach to ethics. However, it now seems to be their critics, with an eye on long-term structural change, who lack appropriate moral concern.

Still, perhaps it is true that more good can be done through political activism than through charity. Perhaps, in other words, you should sometimes let drowning children drown. Utilitarians, for one, are open to counterintuitive moral implications such as this. Recall that MacAskill (2015: 11) says we should "do

what's best, whatever that turns out to be." Suppose, then, for the sake of argument, that suffering in developing countries is the result of structural factors, that structural reform is necessary to alleviate suffering, and that collective political action rather than individual charity has greater priority. Even then, however, the practice might need to be reformed, but EA would yet remain viable. The structural objection, too, doesn't go deep enough to evaluate the theory itself. It speaks only to application of the theory. As Singer (2015b) himself says, "[e]ffective altruism cannot be refuted by evidence that some other strategy will be more effective than the one effective altruists are using, because effective altruists will then adopt that strategy."

EA is the view that we should take a scientific approach to doing good. We should not rely on unreflective judgments about which causes are worthy. Given the moral problems at issue, and given that the solutions are not obvious, we should seek empirical evidence about which solutions are most likely to address them. This principle is broad enough to apply not just to how we spend surplus income but also to how we apportion our political efforts. Thus, some organizations aligned with EA, like The Humane League, focus on structural reform to advance animal welfare (see Weathers 2016). For projects like this, empirical evidence would seem to be necessary. We should not simply devote ourselves to forms of political activism that leave us with a warm glow. We need, in short, not just a science of charity but also a science of political activism. Some philosophers, like Jeff Sebo, have recently taken up the project of studying animal welfare activism (Sebo and Singer 2018; Sebo forthcoming).

To see more clearly how EA can absorb the structural objection, consider the following dilemma. On the one hand, suppose that there is empirical evidence that political activism does the most good, e.g., activism that seeks to reform global economic institutions. This might be measured in terms of quality-adjusted life years (QALYs) or perhaps according to some other measure. Then, by its own lights, EA should accord activism greater priority than charity. On the other hand, suppose that empirical evidence does not show that activism does the most good. Then EA should continue to prioritize charity. But in that case, it seems, any moral agent should think so too. That is, a responsible moral agent should respect scientific evidence about the effects of political activism instead of simply going with their gut. Doing otherwise seems epistemically irrational. As McMahan says, recall, activism vs. charity is not an either/or question. But there are still questions of priority and resource allocation, e.g., the proportion of effective altruists devoted to either activity. EA's science of doing good provides a useful framework for thinking about the social division of moral labor within the movement.

If EA can be reconciled with the structural objection in the way I've suggested, however, further amendments to the usual practice of EA are required. As a theory, effective altruism is committed to a science of doing good. In practice, though, effective altruists tend to be highly selective with respect to scientific evidence. They rely mainly on economics, especially development economics, as relevant to "how to do good better." Economics is well positioned to study the

effects of charity where it is needed most. In practice, then, EA reflects a bias toward interventions that are relatively easy to measure through quantitative methods (Sebo and Singer 2018). In some ways this bias makes good sense. If we can't accurately measure the efficacy of an intervention, we seem to have less reason to support it. Nonetheless, in light of our discussion in this section, the sources of evidence that EA practitioners rely on must be expanded.

We need here to distinguish two types of questions: general and specific. The general question is about what *general causes* most deserve our money or efforts. For example, MacAskill argues that global poverty is one area in which people can do a lot of good. One reason, recall, is that charitable contributions go so much further in poor countries than in the developed world. But MacAskill also thinks other causes are similarly worthy, including factory farming, climate change, immigration, and criminal justice reform (2015: 185-93). Each of these general causes, he argues, involves social problems that are large in scale, neglected, and potentially tractable. More recently, MacAskill has pursued "longtermism," trying to think seriously about how to improve the lives of people thousands or millions of years into the future (supposing that humans make it that far).

General questions about worthy causes lead to more specific questions. Holding fixed the general cause, specific questions remain about *which programs or actions* are the best way of aiding the cause. Thus, for example, MacAskill argues that deworming programs in the developing world have a higher impact on educational outcomes than programs that donate textbooks to schoolchildren (2015: 104-8). Much of MacAskill's book is devoted to using economics to answer various specific questions. When it comes to the area of political activism, however, it's not clear that economics is well positioned to answer specific questions about *how to pursue political activism* (at least given the current state of the discipline). Other scientific fields can potentially offer more insight. For example, work in sociology attempts to study what types of political movements are most effective.

Consider the work of the sociologist Charles Tilly (Tilly 2006; Tilly and Tarrow 2015; see Anderson 2014 for further discussion). Tilly claims that social movements are effective when they can publicly demonstrate the possession of four features: (1) apparent worthiness; (2) unity among the members; (3) high number of adherents; and (4) commitment in the face of personal sacrifice. This type of scientific research is absent from existing EA practice, but it would seem to be crucial if people are to make decisions about how to effectively participate in political activism—not whether to participate in activism at all or what priority it has (general question), but which political activities are most worthwhile (specific question). Without relying on research like this, it seems, effective altruists cannot carry out a sufficiently broad and robust science of doing good.

3. Effective Altruism Without Utilitarianism

I've argued that EA has the ability to meet what is perhaps the most serious criticism leveled against it. However, even if you agree that the structural objection fails in the end, you may still wonder what positive reason there is in favor of EA in the first place. If you think there is no reason to accept EA, you may regard objections and responses as internal disputes that have no wider relevance. Of course, utilitarianism entails EA, but that is no reason to accept EA if there are strong reasons to reject utilitarianism.

In this section, I'll argue that EA need not rest on utilitarianism (see also McMahan 2016; MacAskill 2017). To make this case, I'll continue to rely on an interpretation of EA as a science of doing good. That is, EA is the view that we should rely on scientific evidence to determine how we can do the most good. But I'll argue explicitly now that, given the arguments offered for EA, it does not rule out the existence of other moral obligations, including obligations for which science is not particularly relevant. So, the version of EA that I'll defend does not line up with everything that utilitarians like Singer and MacAskill say about it. However, I'll argue that this makes EA more plausible, not less. The result is a version of EA "worth wanting" and yet one that still imposes rather strong demands on moral agents. Some effective altruists, indeed, might welcome a utilitarianism-free version of EA if it were then, as a consequence, to gain wider appeal and spur more philanthropy and thus be more fruitful on utilitarian grounds.

Let's begin with another persistent objection to EA (aside from the structural objection), one that stems from Bernard Williams' (1973) famous critique of utilitarianism. I won't try to fully unpack Williams' ideas or attempt to be precisely faithful to them, but I hope to say enough to explain why one might still be drawn to EA, or something like it, even if one is sufficiently persuaded by Williams' critique to reject full-blooded utilitarianism.

Williams argues that utilitarianism is unattractive because it is too impersonal and ignores the "separateness of persons" (see Brink, this volume). It treats individuals as mere instruments of utility maximization and ignores their partial perspectives and commitments (see Jeske, this volume). According to utilitarianism, doing the most good is the only thing that matters, no matter the particular goals and projects that a person has and that seem to make her life meaningful. For these reasons, utilitarianism is "alienating."

As McMahan (2016) observes, many critics of EA channel Williams, directly or indirectly attacking EA's allegedly utilitarian foundations. For example, Nakul Krishna (2016) insists that there is value in the supposed "hokeyness involved in the business of finding ourselves and our deepest impulses." Ethics cannot and should not eliminate personal cares from its purview. What one cares about means something:

> [EA presents a] picture of moral reflection as arbitration between the claims of different people, one of whom just happens to be me. In this picture, it seems like the fact that *I'm me* has been declared, right at the outset, irrelevant. To direct my charitable donations to training guide dogs for the blind (an obscenely inefficient way of doing good, the effective altruists say)

> would be to treat (mistakenly) the fact that I happen to care about this cause as if it meant something.

Suppose you are persuaded by the thought that ethics must not treat a person's cares and projects as morally irrelevant. What does this mean for EA? I'll argue presently that it casts doubt on some applications of EA, but that EA itself remains intact.

Williams-style reservations about utilitarianism clash with some of the recommendations that effective altruists make, in particular, those that ignore partiality. MacAskill (2015: 41-42; emphasis added) considers partiality and rejects its moral significance:

> If I were to give to [a foundation with which I have a personal connection] rather than to the charities that I thought were most effective, I would be privileging the needs of some people over others merely because I happened to know them. That would be *unfair* to those I could have helped more…. For example, if an uncle dies of cancer, you might naturally want to raise money for cancer research. Responding to bereavement by trying to make a difference is certainly admirable. But it seems *arbitrary* to raise money for one specific cause of death rather than any other…. By all means, we should harness the sadness we feel at the loss of a loved one in order to make the world a better place. But we should focus that motivation on preventing death and improving lives, rather than preventing death and improving lives in one very specific way. Any other decision would be *unfair* to those whom we could have helped more.

Now, it's possible that many people will have enhanced motivation to donate to charities when they have personal connections to them, possible then that yielding to personal connections will do more good overall than aiming to be most efficient. Were there empirical evidence supporting these ideas, MacAskill and other effective altruists might grant that partiality has moral significance, though only indirect. (On the other hand, motivational dispositions are not to be taken as fixed and can themselves be the target of influence.) However, scrutinizing this example again does not go deep enough to evaluate EA as a philosophical position. Our concern is with the idea that it is "arbitrary" or "unfair" to privilege personal connections over impersonal interests. This can't be convincing to anyone persuaded by Williams (cf. Gabriel 2018).

Utilitarianism rejects the very idea of "special obligations," i.e., obligations that one has not to sentient or sapient beings in general but in virtue of personal relationships. We might have special obligations in virtue of our personal cares and commitments, or in virtue of duties of reparations for past wrongs. Williams gives us reasons to hold on to these obligations. Nonetheless, even if special obligations are not eliminated, any plausible moral theory will have to include so-called "general obligations" that one has to others in general. These include *obligations of beneficence*. Such obligations could but need not be cashed out in terms of happiness. Moreover, it might yet be true that we should fulfill obligations of beneficence through impartial charity or activism, that donating money or time to special causes does not suffice to fulfill these obligations, and

that most of us should donate much more than we currently do. EA thus can still have bite, Williams notwithstanding.

It's striking that in MacAskill's book he nowhere mentions utilitarianism. Nor does he argue for EA on utilitarian grounds, that is, along the lines rehearsed in section 1 of this essay. Like Singer himself, MacAskill makes his case for EA by appealing to moral intuitions that are widely held and seemingly quite plausible. For example, MacAskill (2015: 30-32) describes the process of medical triage and the decisions doctors and nurses must make to prioritize those patients who can most benefit from treatment. Medical professionals must make these difficult decisions, leaving some patients to suffer or die, in order to save others. Decisions about charitable contributions are similar to triage, MacAskill argues. They involve making "hard trade-offs" (32). Or consider MacAskill's (23) argument for channeling financial aid to developing countries, given that money goes roughly 100 times further there than in the developed world:

> It's not often that you have two options, one of which is one hundred times better than the other. Imagine a happy hour where you could buy yourself a beer for five dollars or buy someone else a beer for five cents. If that were the case, we'd probably be pretty generous—next round's on me! But that's effectively the situation we're in all the time. It's like a 99-percent-off sale…. It might be the most amazing deal you'll see in your life.

The strongest argument for EA is thus not that it follows from utilitarianism. And it doesn't lead to a view on which our personal projects are morally insignificant. Rather, EA claims that a science of doing good is the best way to fulfill our obligations of impartial beneficence and that we are required to fulfill these obligations in the best way possible. The strongest argument for this view is that it follows from ordinary, plausible intuitions to which we already seem to be committed, like the intuition Singer evokes in his famous example. Intuitively, one is obligated to save a child drowning in a nearby pond and there is no morally relevant difference in the case of starving children. Intuitively, as well, we should save more children rather than fewer. But it remains an open question whether, in any given case, obligations of beneficence trump special obligations. Without an all-encompassing theory like utilitarianism, such questions can never be closed and will always depend on the details.

I've now argued that EA is best construed as a theory about how to fulfill certain general obligations—in particular, obligations of beneficence—not a consequence of the utilitarian worldview that eschews special obligations. That is, EA says that if we have the time and if available evidence is not already decisive, we should rely on scientific evidence to determine how we can do the most good so as to fulfill obligations of impersonal beneficence. Note that I've simply assumed that there is something to Williams' criticism of utilitarianism, without considering possible responses. The ultimate merits of this criticism, I grant, are not obvious. But my point is that even if you relinquish utilitarianism, there are still powerful reasons to hang on to EA.

Nonetheless, from this perspective, some applications of EA are no longer supported. For example, it may be permissible to give something to charities with

which you have a personal connection. In addition, it does not follow that you should choose a career that enables "earning to give"—not if pursuing a meaningful life also matters. However, EA still offers plenty of other recommendations that a morally responsible agent should act on. A science of doing good is ineliminable if we have a responsibility, as it seems we do, to effectively combat such ills as global poverty, animal suffering, and climate change.

4. Scientific Utilitarianism

Up till now in this chapter, I've been critically reviewing the literature on effective altruism. I've argued that EA is attractive in that it says moral decisions should be grounded in scientific evidence. This virtue of EA can be preserved even if those sympathetic to the view prioritize collective political action over individual financial charity. The virtue can also be preserved even if we abandon a totalizing utilitarianism and maintain a commitment to special obligations and meaning via personal cares and commitments. However, it has been argued by Singer himself, among others, that a scientifically grounded ethics leads, after all, to utilitarianism. In that case, we could avail ourselves of a simpler and more straightforward argument for EA. And the amendment to EA given in the previous section—that it applies only to some of our obligations—would turn out to be unnecessary.

Why would one think that science leads to utilitarianism? A fairly simple-minded reason is that scientists themselves tend to be utilitarians—that is, if they subscribe to any moral theory at all. I'll argue in a moment that, as one might expect, this is not a very good reason for thinking that science supports utilitarianism, not even the best one. But the mere fact that utilitarianism appeals to scientists is interesting enough to merit attention. I'll lay out a couple of explanatory hypotheses for this sociological phenomenon, but I won't spend time defending them since I want to turn to a better scientific argument for utilitarianism, and contend that this argument doesn't hold water either.

One reason many scientists are drawn to utilitarianism might be simply that if offers a mathematical approach to ethical decision-making that is reminiscent of mathematical approaches in science. Utilitarianism asks a moral agent to precisely specify the outcomes of actions, along with their probabilities, in order to calculate the expected value of each possible action. Then, to figure out the morally correct choice, it seems, one simply has to do the math. Thus, a mathematical approach that is sensible in empirical domains may strike scientists as sensible in the ethical domain, too. However, while this might be what causes some scientists to be utilitarians, it isn't a very good reason—not without some argument for thinking that ethics is relevantly like science. Mathematical approaches are not sensible, perhaps not even intelligible, in domains like aesthetics or literature. Furthermore, as Tyler John (personal communication) points out in this context, non-utilitarian and even non-consequentialist moral theories can be formalized too, in ways that can appeal to mathematically-oriented

minds (see Hurley, this volume for relevant discussion of "consequentializing"). As a result, this explanation for utilitarianism's popularity might debunk, rather than vindicate, scientists' commitment to the theory (see Kumar 2017).

I suspect there is an additional reason that scientists are drawn to utilitarianism. The idea of instrumental reasons (or instrumental value) is relatively clear, even to those who lack philosophical training. Instrumental reasons would seem to fit quite naturally within the materialist worldview favored by scientists. To a first approximation, instrumental reasons consist in relations of cause and effect. That something is instrumentally valuable is, or seems to be, just the fact that it helps to bring about a good outcome. Because utilitarianism embodies a minimal commitment to intrinsic or inherent value—it does not truck with rights or justice or any other non-instrumental sources of value or reasons aside from happiness—it appeals to people who are friendly toward instrumental reasons and wary of other normative categories. Herein, I believe, lies an argument that, were it to be unpacked more fully, is potentially quite powerful (which is not to say decisive). That is, scientific materialism might cast doubt on the idea of non-instrumental reasons. A thorough commitment to this argument would lead to moral nihilism rather than utilitarianism. However, even given other reasons to reject nihilism, the argument would support only consequentialism generally and not utilitarianism specifically (see Hubin 2001 for relevant discussion). Whatever it is that has non-instrumental value need not be happiness.

Though they remain only hypotheses, for all that I have said, we have on the table now two explanations for why scientists tend to believe in utilitarianism. However, since these explanations don't vindicate the beliefs (Kumar 2017), we don't yet have an argument for why science supports utilitarianism. We'll turn next to one such argument that is widely discussed in the literature. It is fueled not by economics or sociology but by cognitive science and evolutionary biology.

Utilitarianism is beguiling in its simplicity. It also has quite radical and demanding implications (see Sobel, this volume). Utilitarianism seems to entail that individuals should sacrifice all of their personal interests for the sake of others. Members of the global 1% should not simply funnel some small portion of their resources to the developing world. Given the way in which money goes so much further there, utilitarianism entails that those of us living relatively comfortable lives should donate all of our money until there is no person worse off than us whose interests can yet be advanced. Or, depending on the evidence and the math, we should devote all of our time to collective political action and give up everything else that presumptively makes our lives meaningful.

Many philosophers have thought that utilitarianism is untenable because of other implications that are not just radical but deeply counterintuitive. Utilitarianism denies the existence of rights and justice that transcend happiness, not to mention the moral significance of personal cares and commitments. For these reasons, it seems to violate plain moral commonsense, or even moral decency. However, Singer (2005) claims that although utilitarianism conflicts with widespread moral intuitions, these intuitions are not trustworthy given their evolutionary and psychological origins. Our moral intuitions are the product of the very particular

ecological circumstances that gave rise to the human moral mind (cf. Kumar and Campbell, in prep). As a consequence, Singer (2005: 348) argues, moral commonsense is not to be trusted:

> There is little point in constructing a moral theory designed to match considered moral judgments that themselves stem from our evolved responses to the situations in which we and our ancestors lived during the period of our evolution as social mammals, primates, and finally, human beings. We should, with our current powers of reasoning and our rapidly changing circumstances, be able to do better than that.

Singer is arguing for utilitarianism indirectly, by offering an evolutionary and psychological debunking argument against the intuitions that seem to undermine it. This debunking argument has been developed in more detail by the philosophically-trained psychologist Joshua Greene (2007; 2014a; 2014b), who is also responsible for some of the empirical research that fuels it. Let's focus on his argument.

Greene argues that evolutionary forces gave rise to certain "moral heuristics." These heuristics have a number of cognitive and motivational consequences. They lead us to be partial to our friends and family, since this was crucial in small-scale communities of hunter-gatherers. They also make us averse to harming others through direct "personal force," even when more good can be produced that way, since this aversion was likewise essential to cooperation in the Pleistocene.

In the "environment of evolutionary adaptedness," then, the heuristics that underlie moral intuitions had survival value. Greene suggests that they also likely produced the most good in that environment, relative to feasible alternatives. Nowadays, however, in large scale, technologically-advanced societies, these heuristics lead us astray, according to Greene. Our current environment makes partiality harmful and creates plenty of opportunities through which people can be harmed indirectly without "personal force," including through inaction. Moral intuitions are thus somewhat like the psychological drives that led our ancestors to cash in on rare sources of fat and sugar when times were lean, as they often were in the Pleistocene, but that lead to obesity in many modern environments.

The problem with Greene's arguments, in essence, is that he must rely on the claim that moral intuitions are driven largely by morally irrelevant factors. This claim is not substantiated. For example, Greene does not seem to possess an argument that is independent of utilitarianism, and thus that isn't question-begging, for thinking that partiality is morally irrelevant. Greene's critics should grant that whether or not harm is inflicted through personal force is morally irrelevant. This claim is plausible, but it doesn't go nearly far enough (Kumar 2017: 125-6):

> [Moral] intuitions are sensitive to a range of [other] factors…. For example, intuitions track the degree of harm inflicted, whether it was caused intentionally or only accidentally, whether it was intended as a means to an end or merely as a foreseen side effect, whether it was a deserved response to aggression or unprovoked, and so on…. Greene must [claim] that these

> *other factors* influencing intuition also do not lend rational credibility to…them. The problem for Greene is that this…is not at all plausible.

Greene explicitly does not attempt to leap from "is" to "ought." He offers a debunking argument that rests not just on empirical claims about human psychology but also on normative claims about what is and is not morally relevant. In general, a debunking argument is successful insofar as its normative premises are more plausible than the normative claims it attempts to debunk (Kumar and Campbell 2013). By these standards, Greene's argument is unsuccessful, given the implausibility of the normative claim that all or most of the factors that drive moral intuition are morally irrelevant.

My criticisms of Greene and Singer, like their own arguments, rest on empirical research in cognitive science. More detailed articulation of these criticisms and closer readings of the empirical evidence can be found elsewhere (see Kumar 2017; Kumar and Campbell 2012; Kumar and May 2019). Here I want to end by suggesting a new argument that is in some sense "a priori" in that it doesn't rest on evidence from cognitive science.

Greene argues that the psychological mechanisms that underlie people's moral intuitions are faulty. The rationally innocent mechanisms are those that underlie their (conflicting) commitment to utilitarianism. For the sake of argument, suppose we grant that "the rational parts of ourselves" are drawn to utilitarianism. But for all that cognitive science says, that may be because utilitarianism is deceptively plausible, because it takes in rational minds with its alluring but shallow simplicity. Science can tell us what draws people to utilitarianism, even rationally, but it cannot tell us whether the theory is credible in the end. Only philosophy can do that.

5. Summary

Science doesn't support utilitarianism. But science does play a valuable role in EA. We need a science of doing good no matter what broader ethical theory we subscribe to, even if we subscribe to no broad ethical theory at all. There are two main objections to EA that I've argued do not win the day. If there is good evidence that we can best ameliorate global poverty and suffering through collective political activism that seeks structural reform, then EA should recommend that. If utilitarianism fails to eliminate special obligations that arise from our personal cares and commitments, then EA isn't our only ethical guide but still provides a science of how to fulfill general obligations of beneficence. Singer, MacAskill, and other effective altruists offer persuasive arguments—grounded in intuitions about concrete cases and not in broad ethical theories—that people in developed countries must do more and do better. This isn't all that effective altruists seek to establish, but it is more than good enough.

Acknowledgements

I'm grateful to Samia Hesni, Tyler John, Judith Lichtenberg, Meghan Nesmith, Douglas Portmore, and Aja Watkins for very helpful comments on previous drafts. Work on this chapter was supported by the Peter Paul Professorship at Boston University.

Bibliography:

Anderson, Elizabeth (2014). Social movements, experiments in living, and moral progress: Case studies from Britain's abolition of slavery. The Lindley Lecture, The University of Kansas.

Barnes, Elizabeth (2016). *The Minority Body: A Theory of Disability*. Oxford, UK: Oxford University Press.

Bentham, Jeremy (1789). *An Introduction to the Principles of Morals and Legislation*.

Gabriel, Iason (2018). The problem with yuppie ethics. *Utilitas* 30 (1): 32–53.

Greene, Joshua (2007). "The Secret Joke of Kant's Soul." In *Moral Psychology, Vol. 3*, edited by W. Sinnott-Armstrong. Cambridge, MA: MIT Press.

Greene, Joshua (2014a). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. London, UK: Penguin Books.

Greene, Joshua (2014b). "Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics." *Ethics* 124 (4): 695–726.

Herzog, Lisa (2016). Can 'effective altruism' really change the world? *Open Democracy*, Feb. 22, 2016.

Hubin, Donald (2001). The groundless normativity of instrumental rationality." *Journal of Philosophy* 98(9): 445–468.

Krishna, Nakul (2016). Add you own egg. *The Point Magazine*, January 14, 2016.

Kumar, Victor (2017). Moral vindications. *Cognition* 167: 124–134.

Kumar, Victor, and Campbell, Richmond. On the normative significance of experimental moral psychology. *Philosophical Psychology* 25 (3): 311–330.

Kumar, Victor, and Campbell, Richmond (in prep). *A Better Ape: How Morality Drives Human Evolution*. Unpublished book manuscript.

Kumar, Victor, and May, Joshua (2019). How to debunk moral beliefs. In *Methodology and Moral Philosophy*, edited by Jussi Suikkanen and Antti Kauppinen. Routledge.

MacAskill, William (2015). *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. Gotham Press.

MacAskill, William (2017). Effective altruism: Introduction. *Essays in Philosophy* 18(1).

McMahan, Jeff (2016). Philosophical critiques of effective altruism. *The Philosophers' Magazine* 73: 92–99.

Mill, John Stuart (1863). *Utilitarianism*.

Sebo, Jeff (Forthcoming). Effective Animal Advocacy. *The Routledge Handbook of Animal Ethics*.

Sebo, Jeff, and Singer, Peter (2018). Activism. *Critical Terms for Animal Studies*. University of Chicago Press.

Sidgwick, Henry (1907). *The Methods of Ethics*.

Singer, Peter (1975). *Animal Liberation: The Definitive Classic of the Animal Movement*. New York, NY: Harper Perennial Modern Classics.

Singer, Peter (2005). "Ethics and Intuitions." *Journal of Ethics* 9(3-4): 331–52.

Singer, Peter (1972) Famine, affluence, and morality. *Philosophy and Public Affairs* 1(3): 229–243.

Singer, Peter (2009). *The Life You Can Save: How to Do Your Part to End World Poverty*. New York, NY: Random House Trade Paperbacks.

Singer, Peter (2015a). *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. New Haven, CT: Yale University Press.

Singer, Peter (2015b). The logic of effective altruism. *Boston Review*, July 1, 2015.

Weathers, Scott (2016). Can 'effective altruism' change the world? It already has. *Open Democracy*, Feb. 29, 2016.

Williams, Bernard (1973). A critique of utilitarianism. In Smart, J. J. C., and Bernard Williams, *Utilitarianism for and Against*. Cambridge University Press.

Srinivasan, Amia (2015). Stop the Robot Apocalypse. *London Review of Books*, September 24, 2015.

Taylor, Sunaura (2017). *Beasts of Burden: Animal and Disability Liberation*. New York: The New Press.

Tilly, Charles (2006). *Identities, Boundaries and Social Ties*. Boulder, Colo: Routledge.

Tilly, Charles, and Tarrow, Sidney (2015). *Contentious Politics*. New York, NY: Oxford University Press.