

## How to insure against utilitarian overconfidence

Nicholas Agar

Published online: 13 January 2015  
© Monash University 2015

**Abstract** This paper addresses two examples of overconfident presentations of utilitarian moral conclusions. First, there is Peter Singer’s widely discussed claim that if the consequences of a medical experiment are sufficiently good to justify the use of animals, then we should be prepared to perform the experiment on human beings with equivalent mental capacities. Second, I consider defences of infanticide or after-birth abortion. I do not challenge the soundness of these arguments. Rather, I accuse those who seek to translate these conclusions into moral advice of a dangerous overconfidence. This paper offers an insurance policy that protects against some of the costs of mistaken moral reasoning. An interest in moral insurance is motivated by the recognition that, in the event that overconfident ethicists have reasoned incorrectly, some actions recommended by their conclusions are not just bad, but very bad. We should reject suggestions that we conduct medical experiments on humans or kill newborns

**Keywords** Utilitarianism · Animal experimentation · Infanticide

This paper addresses some dangers of moral overconfidence. I consider two examples of overconfident presentations of utilitarian moral conclusions. First, there is Peter Singer’s widely discussed claim that if the consequences of a medical experiment are sufficiently good to justify the use of animals, then we should be prepared to perform the experiment on human beings with equivalent mental capacities (Singer 1993, pp. 85–89). According to Singer, those who distinguish between animals and humans make a mistake that is no more defensible than racism. Second, I consider defences of infanticide or after-birth abortion. According

---

N. Agar (✉)  
Philosophy Programme, Victoria University of Wellington, P.O. Box 600, Wellington 6140,  
New Zealand  
e-mail: nicholas.agar@vuw.ac.nz

to a recent presentation of this view due to Alberto Giubilini and Francesca Minerva, new-borns, like foetuses, do not possess the attributes that would give them a serious right to life (Giubilini and Minerva 2013). They are not persons. A moral permission to kill foetuses applies also to new-borns.

I do not challenge the soundness of these arguments. Rather, I accuse those who seek to translate these conclusions into moral advice of a dangerous overconfidence. This paper offers an insurance policy that protects against some of the potential costs of mistaken moral reasoning. An interest in moral insurance is motivated by the recognition that, in the event that overconfident ethicists have reasoned incorrectly, some actions recommended by their conclusions are not just bad, but very bad. The defenders of the idea that humans and nonhumans should be equal candidates for medical experiments and of the moral permissibility of infanticide should acknowledge the possibility that they may have reasoned incorrectly. They should grant this possibility even as they dispute the claims of their philosophical opponents.

I draw an analogy between certain moral choices and the decision to insure a house against fire. It can be prudentially rational to insure against the destruction of your house by fire even if you've taken every reasonable precaution to prevent such an event and are confident that your house will not burn down. So too, it can be rational for ethicists to take out moral insurance against performing actions that would be very bad should their moral reasoning be mistaken. This moral insurance policy should lead us to sometimes refuse to act on our moral conclusions.

The moral insurance policy does not direct that we consistently refuse to act in accordance with moral reasoning that we find rationally persuasive. On most occasions, we should act as our moral reasoning directs. A fire insurance policy is worth purchasing when it is cheap relative to the value of your house. A moral insurance policy instructs that we deviate from the dictates of our moral principles only when we recognize that doing so is not costly in moral terms. I offer a utilitarian account of what makes it to cheap to deviate from a moral judgment that you believe to be true. I predict that this utilitarian analysis will be attractive to the overconfident ethicists I discuss. They are either explicit utilitarians or make significant use of utilitarian ideas.

For simplicity's sake, this paper presents moral disputes in cognitivist terms. But what I say is compatible with a non-cognitivist presentation that grants reason a significant role. On one non-cognitivist account we would be disputing not the truth of utilitarianism but instead its rational assertibility.

## **1 Insuring your home against fire and insuring against moral error**

It can be rational to purchase fire insurance even if you have taken every precaution against the destruction of your house by fire. House fires are rare events. But they are both sufficiently probable and sufficiently bad to justify the purchase of an appropriately priced fire insurance policy.

The probability of a house burning down is low but non-negligible. Some bad outcomes with a non-zero probability of occurring are nevertheless insufficiently

probable to justify the purchase of insurance. The destruction of your house by an extra-terrestrial death ray is a logically possible misfortune with a non-zero probability of occurring. There are practical problems for someone who views the non-zero probability of this misfortune as grounds to purchase even a very cheap insurance policy specific to that particular mode of destruction. She will find analogous reasoning compelling her to invest effort considering indefinitely many policies against negligibly probable misfortunes. When considered in the abstract, there is some miniscule fraction of a cent that it is worth spending on a policy against the specific event of house destruction by extra-terrestrial death ray. Home-owners with no limits on their time will find that analogous reasoning should lead them to commit a similarly miniscule amount to insure against destruction by a rogue sea monster, or by a tyrannosaur recently escaped from a cloning lab. Human home-owners with limits on the time that they can commit to contemplating insurance policies will find that the time required to think seriously about what fraction of a cent they should invest in these exotic policies is better spent elsewhere. They are entitled to peremptorily dismiss insurance policies specific to negligibly probable misfortunes. There is therefore an important practical distinction between insuring against misfortunes with non-zero, non-negligible probabilities and insuring against misfortunes with probabilities that are non-zero but negligible.

Now consider a moral case. Suppose that you are a utilitarian. Your understanding of moral theory means that you are, in the terminology of subjective probability, entitled to assign a high credence to the truth of utilitarianism. This has implications for your behaviour. When deciding how you should act, the high credence you assign should direct you to typically act as utilitarianism says. But alternatives to utilitarianism have implications for how appropriately confident utilitarians should act. For example, utilitarians should allow that some intelligent, relevantly informed people who have thought long and hard about ethics arrive at deontological conclusions. Some statements of deontological morality are not demonstrably contradictory. They do not openly conflict with widely acknowledged principles of reasoning. They accord with and potentially explain and justify some of our intuitive judgments about moral rightness. They do not make, or assume, patently false claims about human beings. The same points apply to virtue ethics, contractualist ethics, and almost any view about normative ethics that serious-minded moral philosophers currently defend.

Utilitarians should allow that these are *reasonable* views of ethics, where by “a reasonable view of ethics” I mean a view that we recognize as having a non-negligible probability of being correct. Philosophically conscientious utilitarians may justifiably credit themselves with good reasons for rejecting deontological views whose reasonableness they nevertheless concede. Your purchase of fire insurance does not invalidate your belief that your dwelling will not burn down. You can be confident that your house will be standing when you return from a long summer holiday. So too, a utilitarian’s willingness to acquire moral insurance should not invalidate her belief that utilitarianism is true.

Utilitarians should be discriminating in their selections of alternatives worth insuring against. Some alternatives to utilitarianism should be treated as relevantly

similar to your house's being destroyed by an extra-terrestrial death ray. For example, racist moralities are riddled with inconsistencies and demonstrably false empirical claims. It may be that there is a non-zero probability of some explicitly racist moral theory being true. But everything we know about human beings and morality should lead us to recognize that this probability is nevertheless negligible. Earlier I proposed that there are costs in seriously considering extra-terrestrial death ray insurance. There are costs in taking racist moral arguments seriously. Cognitive resources dedicated to considering arguments for the racial superiority of one group are simply better invested elsewhere—for example in evaluating the moral implications of climate change. A utilitarian who resolves to seriously consider the moral claims of the Aryan Brotherhood so as to appropriately morally insure herself against their possible truth wastes time better put to the consideration of other moral approaches and issues. The same does not appear to be the case for reasonable alternatives to utilitarianism. A utilitarian may justifiably view a colleague who subscribes to virtue ethics as mistaken. But it would be wrong to place virtue ethics in the same category as Nazi morality. It is wrong to assign to virtue ethics a credence that is either zero or so close to zero as to make the view worthy of being ignored. She should consider the implications of failing to act virtuously should virtue ethics be correct. This concession has implications for moral choices. It gives the views of virtue ethicists a relevance to the moral reasoning of utilitarians lacked by the views of racist moralists.

It is rational for utilitarians with an interest in doing what is right to purchase insurance that protects them in the eventuality that their view about moral rightness is mistaken. It is also moral for utilitarians to do so, supposing that their commitment to their utilitarianism permits them to understand that in the (unlikely) event that utilitarianism is false there are actions that are morally good or bad to perform. Their commitment to utilitarianism entails a belief that the theory offers the best account of morality. But accepting that utilitarianism could be false means that they should accept that utilitarianism does not fully capture the content of the term “moral.” They allow that, should utilitarianism turn out to be false, some other account will describe the content of morality. We can therefore make sense of utilitarians sometimes reasoning and acting in explicitly non-utilitarian ways.

Under what circumstances should appropriately confident utilitarians reject the advice of their moral theory? The circumstances under which it is right to reject utilitarian directives are quite rare. A worthwhile fire insurance policy should be cheap relative to the value of your house. A utilitarian's insurance policy against the falsehood of utilitarianism should be cheap in utilitarian terms. It achieves this cheapness by allowing that insured utilitarians most often act in accordance with the theory that they have good reason to accept. It directs them to reject utilitarian advice only in circumstances in which doing so is a combination of very bad in non-utilitarian terms, and not very bad in utilitarian terms. The cost of deviations from utilitarianism can be assessed in the following way.

**Pricing Moral Insurance for Utilitarians:** For each choice on which utilitarianism differs from reasonable alternatives, the cost for utilitarians of deviating from their moral theory is the difference in utilitarian terms of doing

as utilitarianism directs and acting in accordance with the requirements of other reasonable moral views.

This proposal groups together reasonable alternatives to utilitarianism. There are, of course, many differences between reasonable non-utilitarian moralities. The directives of Kantianism differ in quite significant ways from the recommendations of virtue ethics. However, in some moral debates, utilitarian conclusions are properly recognised as outliers among reasonable views. This is certainly the case in the two debates I discuss in this paper.

Utilitarianism has a scalar theory of value that makes it especially amenable to quantifying the badness of moral errors. Utilitarians can make sense of an act's being morally wrong, but nevertheless not very morally wrong when compared with other moral wrongs. Deliberately killing five persons is less bad for utilitarians than deliberately killing ten persons. Deliberately causing 5 min of suffering to a sentient being is not as bad as deliberately causing 10 min of suffering of the same intensity. Toward the low end of this spectrum of utilitarian moral wrongs we have moral costs analogous to the cheap fire insurance policy.

Other theories do not have scalar theories of value. For example, Kantians view certain acts as morally forbidden. They say that those acts should not be performed. It makes little sense for Kantians to say that a comparison of two morally forbidden acts will reveal that one is *more* morally forbidden than the other. An act is either morally forbidden or it is not. This may make the approach I describe in this paper a less suitable tool for Kantians who seek to minimize the costs of mistaken moral reasoning. It is less easy for Kantians to make sense of moral insurance having a price that is cheap in what they acknowledge as moral terms.

This utilitarian account of the cost of moral insurance should appeal to the overconfident ethicists I discuss. Singer is a utilitarian. Giubilini and Minerva make use of utilitarian ideas. Their defence of after-birth abortion requires no endorsement of the principle of utility. But it does require the rejection of some of the non-utilitarian ideas that would stand in the way of the above proposal about how to price moral insurance.

In what follows I argue that morally insured ethicists should distinguish between medical experiments on humans and nonhumans even when the moral theory that they endorse fails to find any difference between animals properly considered to be good candidates for experimentation and humans with equivalent mental capacities. I argue that morally insured ethicists who endorse abortion should reject the killing of new-borns even when their preferred moral theory finds no difference between foetuses and new-borns.

## **2 Why those prepared to conduct medical experiments on animals should not conduct them on mentally disabled humans**

Singer argues that we have no legitimate moral grounds for distinguishing between nonhuman animals who fail to meet the criteria for personhood and mentally disabled humans who fall equally short of these criteria. He accuses those who

claim to find a moral distinction where there actually is none of speciesism—the conflation of the boundaries of biological species with distinctions in moral worth.

Suppose that it's true that utilitarianism finds no moral difference between a severely mentally handicapped human and a nonhuman animal with equivalent cognitive capacities. Utilitarians should acknowledge that their theory disagrees with many other reasonable moral views in finding no moral difference between members of the human species and members of other species. A severely mentally handicapped human has ties of kinship to other humans. He or she is the son, daughter, sister, or brother of human persons. These ties are the bases of relationships that non-utilitarian normative principles place great importance on. Virtue ethicists would describe as deplorable a character trait that permitted no preference for fellow humans. These alternative accounts of the moral significance of the boundary between humans and nonhumans have a non-negligible probability of being true.

Consider these differences in light of a debate that took place between Singer and Tipu Aziz, a prominent British medical researcher (“Monkeys, Rats, and Me: Animal Testing” BBC documentary). Singer was asked whether he would accept that it was right to inflict suffering on rhesus monkeys if doing so would bring closer a cure for Parkinson’s disease. Singer allowed that it could be. This concession was viewed by some media commentators as a significant reversal by the philosopher famous as the author of numerous works defending the moral standing of animals (Neale 2006). Singer was quick to point out that his concession was not a slip, but rather a simple implication of his utilitarianism. In a letter to the *Sunday Times* newspaper Singer patiently explained “Since I judge actions by their consequences, I have never said that no experiment on an animal can ever be justified... If an experiment on a small number of animals can cure a disease that affects tens of thousands, it could be justifiable” (Singer 2006).

Singer’s complaint is not that Aziz is prepared to inflict suffering on rhesus monkeys to find a better treatment for a very nasty neurological disorder. Rather it’s about what he presumes Aziz would not do. Singer says, “In my book *Animal Liberation* I propose asking experimenters who use animals if they would be prepared to carry out their experiments on human beings at a similar mental level—say, those born with irreversible brain damage. I wonder if Professor Aziz would declare whether he considers such experiments justifiable. If he does not, perhaps he would explain why he thinks that benefits to a large number of human beings can outweigh harming animals, but cannot outweigh inflicting similar harm on humans.” Aziz stands accused of speciesism, a morally illegitimate privileging of human interests over the similar interests of nonhumans. Singer continues that “a prejudice against taking the interests of beings seriously merely because they are not members of our species is no more defensible than similar prejudices based on race or sex” (Singer 2006).

A moral insurance policy that grants alternatives to Singer’s view a non-negligible probability of being true should grant that reasonable moralities condemn a decision to conduct on humans experiments of the type that Aziz performs on rhesus monkeys. Suppose that the experiments are justified in terms of the possible benefits that they bring to people with Parkinson’s disease. The choice is over

whether these experiments should be conducted on rhesus monkeys or on humans with equivalent or inferior mental capacities. Utilitarians should allow that comparatively little is lost in refraining from experimenting on a human being with cognitive capacities similar to those of a rhesus monkey. There are some costs. Humans are likely, in general, to be better experimental models of human disease. An experimental therapy that reverses Parkinson's in a human experimental subject may be more likely to reverse the disease in human patients than is an experimental therapy that reverses the disease in a rhesus monkey. But these costs seem comparatively small when compared with the moral costs avoided by successful banning of experiments on human subjects. Aziz avoids acting in a way that would be very wrong according to some reasonable non-utilitarian views while causing little additional suffering.

Note that this is not a philosophical argument against Singer's claims about the moral irrelevance of species boundaries. It challenges no premise in Singer's argument. It advances no alternative view of moral considerability. Rather it assumes that there are reasonable views of morality that reject Singer's conclusion. Further, it assumes that we should accept that there is a real possibility that one of these alternative views is correct. Some of the views that reject Singer's conclusion are moral theoretic analogues of house destruction by extra-terrestrial death ray. But others are properly compared with destruction by fire.

Suppose future moral discoveries made us certain of utilitarianism's truth. Suppose also that Singer is correct about what utilitarianism requires. In such circumstances privileging humans over nonhumans with similar cognitive powers would be indefensible. We should do exactly as Singer says. These imagined future circumstances are very far from our own, however. In our current circumstances utilitarians do best by sometimes acting as their moral theory tells them they should not.

What should this approach make of Singer's comparison of speciesism with racism? Suppose that, when expressed in utilitarian terms, appeals to species membership are as unfounded as are moral appeals to racial categories. We can nevertheless distinguish these moral appeals in terms of their reasonableness. There are no reasonable moral arguments that make members of one racial group more worthy than members of other racial groups. Deontological, virtue ethical, and contractualist appeals to the moral significance of human attributes or relations with humans should be taken more seriously than racist moral arguments.

### **3 Why those prepared to abort fetuses should refuse to kill new-borns**

Giubilini and Minerva argue that it is morally permissible to kill new-borns. They prefer the term "after-birth abortion" to the more traditional "infanticide" because it emphasises "that the moral status of the individual killed is comparable with that of a fetus ... rather than to that of a child" (Giubilini and Minerva 2013, pp. 261–262). The permission defended by Giubilini and Minerva applies both to disabled and non-disabled new-borns. Their argument echoes earlier defences of infanticide due to Michael Tooley (1972) and Singer (1993). New-borns resemble

foetuses in lacking the attributes required to be moral persons. After-birth abortion prevents a new-born from acquiring personhood, but doing so does not constitute a harm.

The paper prompted a storm of protest. I propose to consider these protests as evidence of the existence of reasonable moral views strongly opposed to after-birth abortion. Let's grant that Giubilini and Minerva can defend their position against these challenges. We can still accuse them of overconfidence in seeking to translate their philosophical conclusion into practical advice. They neglect the need for moral insurance that protects them against the possibility of error.

I have recommended that we take out moral insurance that we acknowledge as cheap to protect us against performing acts that would be very morally bad should other reasonable moral views be true. There are moral costs in preventing reluctant parents of new-borns from performing acts that we may have good reason to believe are morally justified. In many societies unwanted children can be put up for adoption. Giubilini and Minerva make the point that this option is not cost-free. "Birthmothers are often reported to experience serious psychological problems due to the inability to elaborate their loss and to cope with their grief. It is true that grief and sense of loss may accompany both abortion and after-birth abortion as well as adoption, but we cannot assume that for the birthmother the latter is the least traumatic" (Giubilini and Minerva 2013, p. 263). We should acknowledge that these are genuine costs. But the cost of moral insurance here seems quite small. We can certainly take steps to mitigate the trauma of parents who suffer because they know that their biological child is alive somewhere out there. When viewed as a way of avoiding acts that, according to many reasonable views are not just wrong but very wrong, these costs are small.

#### 4 Two objections

It might be objected that this paper's proposal is unjustifiably conservative. It is biased against controversial moral claims. This bias is appropriate. What's problematic about the claims of Singer and Giubilini and Minerva is not that they are controversial, i.e. false according to some reasonable views about morality. It's that they could be quite seriously mistaken. The authors seem quite unconcerned about the possibility that their moral advice might be not only wrong, but very wrong. It's one thing to speculate about the possible moral defensibility of experimenting on the mentally disabled and killing babies. It's another to expect people to act on these speculations. An analogy is instructive. Financial advisers may enjoy speculating about get-rich-quick schemes. But they should be careful before telling others to invest their retirement savings in them. It can be intellectually stimulating to reflect on possible lines of philosophical support for unconventional moral claims. But those who engage in this speculation who are aware of the costs of error should be careful not to actually recommend such acts.

This paper's moral conservatism should not be confused with a political conservatism that holds that we should tolerate conditions that we believe to be immoral. It suggests that we focus our efforts on the many problems apparent to



advocates of a wide range of reasonable views. We should not spend too much time on issues whose appeal is limited to those with moral interests best described as eclectic.

Perhaps this paper assumes too literal an interpretation of the claims of Singer and Giubilini and Minerva. They may have no expectation that we will perform the acts that they recommend. Singer does not really want researchers performing experiments that are justified in utilitarian terms to be as willing to perform these experiments on disabled humans as they are on rhesus monkeys. Giubilini and Minerva do not really want people to conclude that they are morally entitled to kill their babies and to act accordingly. They are content to have those who conduct medical experiments or perform abortions reflect on the moral qualities of their actions. They want supporters of these acts to consider possible parallels between the actions that they take to be morally justified and actions that they find morally repellent.

If this is so, then these authors should be more circumspect in their uses of moral language. We can suppose that when Singer argues that we are morally required to make sacrifices to improve the situations of people experiencing extreme poverty that he intends us to act accordingly (Singer 1972). He is not content merely to offer affluent people grounds for reflection on how they stand in relation to the poor. These uses of moral language should be distinguished from his criticisms of Tipu Aziz and others who conduct medical research on animals. Here Singer would be, on this interpretation, content merely to offer grounds for moral reflection. I suspect that this is a distinction in the use of moral language that Singer would be unwilling to make.

## 5 Concluding comments

This paper has challenged no premise in the argument of any overconfident ethicist. Rather, it has been concerned to offer a way to manage moral epistemic risk. When ethicists arrive at controversial conclusions they should acknowledge that they may have reasoned in error. They should interpret widespread opposition as evidence for reasonable ethical views that reject the controversial conclusion. This should not deter them from presenting their arguments. But it should motivate them to make clear that there is a gap between the kinds of claims that it is interesting for philosophers to advance and the kinds of actions that ethicists should recommend. Appropriately insured ethicists will not recommend that medical experiments be performed on humans or that infants be killed, even when such claims are suggested by moral theories that they have good reason to endorse.

**Acknowledgments** I am grateful to audiences in Mexico City and at Victoria University of Wellington and anonymous referees for this journal.

## References

- Giubilini, A., and F. Minerva. 2013. After-birth abortion: Why should the baby live? *Journal of Medical Ethics* 39: 261–263.

- Monkeys, Rats, and Me: Animal Testing documentary broadcast on the BBC 2006—<http://www.imdb.com/title/tt3132130/>.
- Neale, G. 2006. Peter Singer: Monkey business. The *Independent* 03 December 2006. Available at <http://www.independent.co.uk/news/people/profiles/peter-singer-monkey-business-426768.html>.
- Singer, P. 1972. Famine, Affluence, and Morality. *Philosophy & Public Affairs* 1: 229–243.
- Singer, P. 2006. Letter to the editor of the Sunday Times, <http://www.utilitarian.net/singer/by/20061203.htm>. For further discussion, see the FAQ linked from Singer's homepage, <http://www.princeton.edu/~psinger/faq.html>.
- Singer, P. 1993. *Practical ethics*, 2nd ed. Cambridge: Cambridge University Press.
- Tooley, M. 1972. Abortion and infanticide. *Philosophy & Public Affairs* 1: 37–65.