

## Is the 'trolley problem' a problem?

R Glynn Owens DPhil(Oxon)

The recent death of Oxford philosopher and ethicist Philippa Foot marks a sad loss to the intellectual community. It does, however, provide a timely reminder of the problem for which she is perhaps most famous—the so-called 'trolley problem'—and an opportunity to reconsider the implications of this both for medical ethics and for moral philosophy. The trolley problem has been variously described, but basically asks us to consider the scenario of a runaway rail truck hurtling along the rails and unstoppable. Where we are standing the railway forks; if the truck continues along its present path, it will hit and kill a group of five people tied to the track. If, however, we switch the points, it will go along the other path, where it will hit and kill one person similarly tied. When asked whether we should switch the point to ensure that one person rather than five dies, most people agree that this is the more desirable option. In a parallel example, people are presented with a scenario in which the truck will pass under a bridge, after which it will kill five people; if however a heavy weight is dropped in front of the truck they will be saved. Typically when people are asked to imagine themselves standing on the bridge and in a position

to push a heavy person standing next to them into the path, they are much more reluctant to agree that this is acceptable, even though numerically the examples are equivalent.

The problem has a particular relevance to the world of medicine because of its extensive development by the American philosopher Judith Jarvis Thomson. In a version of the problem she conceived a scenario in which a doctor has five patients in need, respectively, of transplanted kidneys, heart, lungs and liver. The doctor then has another patient turn up for a routine assessment who just happens to be a perfect match for the other five patients. Is the doctor justified in saving the lives of the other five by removing all the relevant organs from the healthy person? Not surprisingly, despite the apparent logical equivalence with previous illustrations, most people argue that such a course of action would be completely unacceptable.<sup>1</sup>

This broadly defined problem has had a marked influence in the world of ethics. It has been seen as a conclusive counterexample to a utilitarian approach, since the latter, it is argued, would unequivocally support the sacrificing

of one patient to save five; it has also been seen as evidence for a moral divide between active and passive routes to a particular outcome. Experimental studies of people's judgements have led, as a result of considerable consistency across a range of variables, to the notion of an 'unconscious moral grammar',<sup>2</sup> with attempts even being made by neuropsychologists to link such moral judgements to particular patterns of neural activity in fMRI studies (but see Uttal<sup>3</sup> for a critique of attempts at localising such functions through fMRI work).

It is, of course, necessary to exercise a certain willing suspension of disbelief in considering these problems, especially given their inherent implausibility. Thus Thomson's version, it quickly becomes apparent, is not a true parallel to Foot's original. If Thomson's five sick and one healthy patient had indeed all the same antigenic status, then the problem simply reduces to one of whether or not it would be justifiable to remove one kidney from the healthy individual. If this is done, then the recipient has (the practicalities and limitations of transplantation aside) the same status as the original putative donor, and would be an equally good candidate for the donation

### CORRESPONDENCE TO:

R Glynn Owens

Professor of Psychology,  
The University of  
Auckland, PB 92019,  
Auckland, New Zealand  
g.owens@auckland.ac.nz

The **ETHICS** column explores issues around practising ethically in primary health care and aims to encourage thoughtfulness about ethical dilemmas that we may face.

**THIS ISSUE:** Psychology professor Glynn Owens takes a relativistic and naturalistic stance on moral judgements. He argues eloquently that ethics are not absolute but vary according to societal contexts. Cultures survive if they are wise in the moral decisions they choose to reward.

of the remaining organs. By extension, any further transplantation similarly cycles the potential donors. At a simpler level, the arrival of the healthy patient for a check-up becomes pretty much an irrelevance, since any of the sick patients could already have been donors for the others, and any one chosen at random could save the rest (or most of the rest, if the donor is the one with the failed kidneys).

Fortunately for the philosophers, such objections are not seen to carry much weight, since a further hypothetical condition can be introduced such that the immunological status is specified to be non-commutative, i.e. person A can be donor to B but not vice versa (as might occur with ABO blood grouping, for example). In such a form, it is argued, most people are uncomfortable with the healthy individual being sacrificed for the greater good.

Perhaps what this example demonstrates more than anything, however, is not the viability or otherwise of utilitarianism, or the failure of logic in making moral decisions, but rather the structure of the basic principle used by moral philosophers. Irrespective of the particular ethical strategy being promulgated, the general approach tends to have something of the following form. Philosopher A comes up with a strategy which will solve moral problems (e.g. Mill's Utilitarianism). Philosopher B then comes up with an example where use of A's strategy leads to a counter-intuitive conclusion (e.g. Foot's trolley problem). On seeing this, the supporters of Philosopher A's approach throw up their hands in despair on realising that their strategy does not work after all, and return to the drawing board.

A second point, perhaps of especial significance in the medical context, is that for the philosopher there is no problem regarding the testing of ethical

strategies against purely hypothetical scenarios. Should a critic object that it is statistically inconceivable that a single doctor would have five sick patients and a healthy one all with close tissue matches, or (as above) that if they were so close that any with two functioning kidneys would be able to save all the others without recourse to the healthy person, this would be shrugged off. 'Just imagine that for some reason the organs all have to come from the healthy person' would be the likely response 'and consider what would be right in such a situation'.

Thus, moral philosophy is commonly characterised by (a) a rejection of any strategy which would direct us to intuitively unacceptable courses of action and (b) a willingness to invoke scenarios

those of us with a scientific background, fairly simple. In no area of science is the fact that a theory or model leads to counterintuitive predictions seen as damning (note that this is not the same as 'thought experiments', which may indeed show unacceptability—but through demonstrating that a stance leads to self-contradiction, not simply a surprising result). When Einstein first published his special and general theories of relativity, a number of counterintuitive implications were apparent. That two people may be the same age on earth, but different ages once one had been sent at high speed into space and back, makes no intuitive sense—but the slowing of time as one approaches the speed of light has been empirically demonstrated, and no-one rejects relativity theory simply on the grounds that

**Unless we are willing to anticipate a future in which all agree on what is right and wrong, and all continue to do so, we can never achieve an ethical strategy which will meet the conditions set by moral philosophers**

which are purely hypothetical in order to provide challenges to such strategies. For those of us grounded in empirical science these are perhaps rather alien notions, and it is a little surprising that they are so rarely questioned. The lack of concern with whether or not a scenario is real or hypothetical fits rather poorly with those of us who have to deal with the real world, and we could perhaps be forgiven for being happy to go along with any ethical strategy whose limitations exist only in hypothetical, not real situations. It is, however, the fundamental methodology of traditional moral philosophy that I would most like to address. Whilst appealing at first glance, at least two objections can be raised against it. The first is, for

the results are counterintuitive.<sup>4</sup> Rather, we acknowledge that at times our intuition is simply wrong. Perhaps we should also accept that in moral judgements our intuition may be similarly misleading?

This observation brings us to the second objection to the traditional approach. If we are to reject any ethical theory which leads to counterintuitive conclusions, then the only theory which could ever be acceptable would be one consistent with everyone's intuition. Yet a moment's reflection suggests that this is likely to remain a holy grail, forever out of reach, since it could only be achieved if everyone's intuition led them to similar conclusions. Even the philosophers have long noted the wide varia-

tion across societies, times and cultures in what is or is not acceptable. Unless we are willing to anticipate a future in which all agree on what is right and wrong, and all continue to do so, we can never achieve an ethical strategy which will meet the conditions set by moral philosophers—and even if such a society were to evolve, the question would remain as to whether their agreed morality could be seen as ‘right’ by comparison with the views (of some, at least) of their ancestors.

So where does this take us? At the very least, I would argue, it suggests that the traditional approach adopted by philosophers is inherently sterile. Arguably, moreover, it suggests that if we are ever to devise ethical principles which are of practical value, we should

One of the incidental ways in which the trolley problem (where we started) has taken moral philosophy forward has been the way in which it has taken the subject into the empirical realm, with a large number of studies examining how judgements might vary as a function of subtle manipulation of parameters. One might go even further, and argue that it is timely to attempt, at least, to understand moral problems in terms of empirical science rather than abstract philosophy. Early attempts such as Spencer’s to reconcile moral philosophy with Darwinian biology made a bad start, with a simplistic notion of ‘survival of the fittest’. Such an approach is neither wholly consistent with Darwinism, nor acceptable. But, more recently, psychologists such as Skinner<sup>5</sup> and ethicists such as Hocutt<sup>6</sup> have

are wise in what they choose to reward will survive. From such a perspective issues of right and wrong, of course, also become dynamic—what is right for a society at one point in time may not be right at another. Thus, in a society which does not understand the role of vitamin B12, declaring that to follow a vegan diet is ‘good’ will lead to the death of the society; in a society which appreciates that the vitamin is essential, a vegan diet may be seen as acceptable with suitable precautions, and may even be seen as ‘good’ if other food resources are in short supply or if animal farming is seen to be environmentally detrimental. In such terms, moral judgements may be seen as relativistic, in the sense that what is good now is not necessarily what was good 500 years ago, and naturalistic, in the sense that eventu-

**What is right, therefore, is what a society chooses to encourage—and over time only those cultures which are wise in what they choose to reward will survive. From such a perspective issues of right and wrong, of course, also become dynamic—what is right for a society at one point in time may not be right at another**

exercise some caution about relying on hypothetical scenarios to test them. To some extent, of course, we recognise this already, and today few medical students would qualify without encountering the pragmatic quadumvirate of beneficence, non-maleficence, autonomy and justice. But even if we restrict ourselves to the real world, these guiding principles may serve either to be inadequate (determining a just course of action may be far from simple—for example, in palliative care, is our prime responsibility to the dying patient or to loved ones who will be left behind to grieve?) or conflicting (autonomy may argue for euthanasia, or the supplying of performance-enhancing drugs to athletes, but non-maleficence would argue against).

suggested that the general principle of natural selection may not only provide a clearer understanding of ethics but also resolve old debates about naturalistic versus relativistic ethics. In their perspective, what is ‘right’ or ‘wrong’ are empirical statements about particular societies. What a society sees as right is what that society or culture chooses to reward. From this perspective certain courses of action are rewarded by particular cultures; different cultures will choose to reward different actions, and over time some will survive and some will fail (a prime example is perhaps Hitler’s ‘Thousand Year Reich’, which in practice survived less than two decades). What is right, therefore, is what a society chooses to encourage—and over time only those cultures which

ally survival will judge whether or not our culture chose to reward the right things. Perhaps most encouragingly of all, we won’t need to be distracted by hypothetical scenarios.

## References

1. Thomson J. Killing, letting die and the trolley problem. *Monist*. 1976;59:204–217.
2. Mikhail J. Universal moral grammar; theory, evidence and the future. *Trends Cogn Sci*. 2007;11:143–152.
3. Uttal W. Hypothetical higher-level cognitive functions cannot be localized in the brain; another argument for a revitalized behaviourism. *Behav Anal*. 2004;27:1–6.
4. Feynman, R. The Feynman lectures on physics Vol. 1. Reading, Massachusetts: Addison-Wesley; 1953.
5. Skinner BF. About behaviorism. London: Jonathan Cape; 1974.
6. Hocutt M. Skinner on the word ‘good’: a naturalistic semantics for ethics. *Ethics*. 1977;87:319–338.