Cognition 134 (2015) 193–209

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good

Guy Kahane [a,b,1], Jim A.C. Everett [a,c,*,1], Brian D. Earp [a], Miguel Farias [d], Julian Savulescu [a,b]

[a] Oxford Uehiro Centre for Practical Ethics, University of Oxford, Littlegate House, St. Ebbe's St., Oxford OX1 1PT, UK
[b] Oxford Centre for Neuroethics, University of Oxford, Littlegate House, St. Ebbe's Street, Oxford OX1 1PT, UK
[c] Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK
[d] Centre for Research in Psychology, Behaviour & Achievement, Coventry University, Priory Street, Coventry CV1 5FB, UK

ARTICLE INFO

Article history:
Received 9 May 2013
Revised 13 October 2014
Accepted 14 October 2014

Keywords:
Moral judgment
Moral dilemmas
Utilitarianism
Psychopathy
Altruism
Impartiality

ABSTRACT

A growing body of research has focused on so-called 'utilitarian' judgments in moral dilemmas in which participants have to choose whether to sacrifice one person in order to save the lives of a greater number. However, the relation between such 'utilitarian' judgments and genuine utilitarian impartial concern for the greater good remains unclear. Across four studies, we investigated the relationship between 'utilitarian' judgment in such sacrificial dilemmas and a range of traits, attitudes, judgments and behaviors that either reflect or reject an impartial concern for the greater good of all. In Study 1, we found that rates of 'utilitarian' judgment were associated with a broadly immoral outlook concerning clear ethical transgressions in a business context, as well as with sub-clinical psychopathy. In Study 2, we found that 'utilitarian' judgment was associated with greater endorsement of rational egoism, less donation of money to a charity, and less identification with the whole of humanity, a core feature of classical utilitarianism. In Studies 3 and 4, we found no association between 'utilitarian' judgments in sacrificial dilemmas and characteristic utilitarian judgments relating to assistance to distant people in need, self-sacrifice and impartiality, even when the utilitarian justification for these judgments was made explicit and unequivocal. This lack of association remained even when we controlled for the anti-social element in 'utilitarian' judgment. Taken together, these results suggest that there is very little relation between sacrificial judgments in the hypothetical dilemmas that dominate current research, and a genuine utilitarian approach to ethics.

1. Introduction

According to classical utilitarianism, we should always aim to maximize aggregate welfare (Bentham, 1789/1961; Mill, 1861). Utilitarianism is a radically impartial view: it tells us to consider things as if 'from the point of view of the universe' (Sidgwick, 1907), without giving any special priority to ourselves, or to those dear or near to us. Instead, we should transcend our narrow, natural sympathies and aim to promote the greater good of humanity as a whole, or even the good of all sentient beings (Singer, 1979). Needless to say, this view of morality is strongly at odds with traditional ethical views and common intuitions. It is also a highly demanding moral view, requiring us, on some views, to make very great personal sacrifices, such as giving most of our income to help needy strangers in distant countries (Kagan, 1989; Singer, 1972).

A great deal of recent research has focused on hypothetical moral dilemmas in which participants must decide whether to sacrifice the life of one person in order to save

* Corresponding author at: Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.cognition.2014.10.005
0010-0277/

the lives of a greater number. In this large and growing literature, when individuals endorse this specific type of harm they are described (following Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) as making *utilitarian* judgments; when they reject it, they are said to be making *non-utilitarian* (or deontological) judgments.[2] This terminology suggests that such 'utilitarian' judgments express the kind of general impartial concern for the greater good that is at the heart of utilitarian ethics. This is a widely held assumption. For example, it has been argued that this research shows that utilitarian judgment is uniquely based in deliberative processing involving a cost-benefit analysis of the act that would lead to the greatest good, while, by contrast, non-utilitarian judgment is driven by instinctual emotional aversion to causing 'up-close-and-personal' harm to another person (Greene, 2008). It has even been argued that this empirical evidence about the psychological sources of utilitarian and non-utilitarian judgment can help explain the historical debate between utilitarians and their opponents (Greene, Nystrom, Engell, Darley, & Cohen, 2004) and, more radically, even that it should lead us to adopt a utilitarian approach to ethics (Greene, 2008; Singer, 2005).

However, as we have pointed out in earlier work, these large theoretical claims are problematic. This is because endorsing harm in the unusual context of sacrificial dilemmas need not express anything resembling an impartial concern for the greater good (Kahane, 2014; Kahane & Shackel, 2010). Indeed, the sacrificial dilemmas typically used in current research represent only one, rather special, context in which utilitarian considerations happen to directly conflict with non-utilitarian rules or intuitions. To be willing to sacrifice one person to save a greater number is merely to reject (or overrule) one such non-utilitarian rule. Such rejection, however, is compatible with accepting extreme non-utilitarian rules in many other contexts—rules about lying, retribution, fairness or property, to name just a few examples, not to mention non-impartial moral norms permitting us give priority to ourselves, and to our family or compatriots, over others. Indeed, to reject a specific non-utilitarian moral rule (or even many such rules) is not yet to endorse the core of utilitarianism: the positive aim of impartially maximizing the greater good of all.

It therefore cannot be assumed that a tendency to make 'utilitarian' judgments in sacrificial 'personal' dilemmas really reflects any kind of genuine concern for the greater good. In fact, two recent studies observed no correlation or even a negative correlation between a tendency to make such 'utilitarian' judgments and seemingly genuine utilitarian judgments or attitudes in other contexts. First, in a prior study, we found no correlation between rates of 'utilitarian' judgment and utilitarian views in a context in which utilitarian considerations were pitted against rules against lying or disrespecting autonomy (Kahane et al., 2012). Second, clinical populations have been reported to exhibit both higher rates of 'utilitarian' judgment in personal moral dilemmas (Koenigs et al., 2007) as well as greater rates of punitive responses to unfair offers in the Ultimatum Game (Koenigs & Tranel, 2007)—retributive responses that are at odds with a strict utilitarian cost-benefit analysis. A 'utilitarian' bias in the context of sacrificial dilemmas thus may not carry over to other contexts, casting doubt on the assumption that it is driven by a general concern with maximizing the good.

Even more strikingly, several recent studies found that 'utilitarian' judgment is associated with *anti*-social traits such as psychopathy (Bartels & Pizarro, 2011; Glenn, Koleva, Iyer, Graham, & Ditto, 2010; Koenigs, Kruepke, Zeier, & Newman, 2012; Wiech et al., 2013), as well as with diminished empathic concern (Choe & Min, 2011; Crockett, Clark, Hauser, & Robbins, 2010). It seems rather implausible that individuals with antisocial traits or lower levels of empathy are especially morally committed to promoting the greater good, or harbor a special concern for humanity as a whole.

Suggestive as this recent evidence may be, the relationship between 'utilitarian' judgment in sacrificial dilemmas and impartial utilitarian concern for the greater good has not yet been examined in a direct and robust fashion. It cannot be ruled out, for example, that some individuals with lower empathy may nevertheless arrive, in a 'cold' fashion, at a more general utilitarian outlook. Moreover, even if there is an antisocial component driving some 'utilitarian' judgments, it remains possible that, once this component has been controlled for, a pattern strongly associating 'utilitarian' judgment and general concern for the greater good will emerge.

The aim of the present study was therefore to directly investigate the relation between 'utilitarian' judgment in sacrificial dilemmas and clear markers of impartial concern for the greater good in other moral contexts (e.g. increased altruist concern for distant strangers) and within the context of sacrificial dilemmas (e.g. willingness to sacrifice oneself to save a greater number), as well as their contraries (e.g. support for egoism or greater willingness to sacrifice someone when this also benefits oneself).

Now, if a strong tendency toward 'utilitarian' judgment in classical sacrificial dilemmas really reflects giving greater (or even exclusive) priority to impartial promotion of the good of all, or a preference for a utilitarian style of moral reasoning—as implied by much of the current work in this area—then we should expect this tendency to be observable as well in *other* contexts in which impartial utilitarian concern for the greater good competes with self-interest and with other moral concerns.

In contrast, if a tendency to 'utilitarian' judgment reflects a narrower moral disposition largely driven, not by concern for the greater good, but by reduced aversion to harming others (Crockett et al., 2010; Cushman, Gray, Gaffey, & Mendes, 2012), then we would expect *no* association between a 'utilitarian' bias in this special context and greater endorsement of paradigmatic utilitarian judgments in other contexts. Moreover, to the extent that such a

[2] There is more than one form of utilitarianism. However, the literature on utilitarian judgment in current moral psychology assumes a simple form of what philosophers call Act Utilitarianism (see e.g. Cushman, Young, & Greene, 2010; Greene, 2008)—a view broadly similar to that associated with utilitarians like Peter Singer. When we refer to utilitarianism, we shall therefore mean only something like this view. In fact on some other forms of utilitarianism (e.g. rule utilitarianism) it would not be obvious that one should, e.g., push an innocent person to his death in order to save a greater number (see Kahane & Shackel, 2010).