Utilitarian theorists observe that humans have two motivations in life, what Bentham (2000, p. 14) describes as “two sovereign masters”: pursuit of happiness, and avoidance of pain. Utilitarian theory assumes that happiness is the ultimate ‘good’, and other goods are whatever brings greater happiness. Likewise, pain is the ultimate ‘bad’ and should be avoided. They suggest that this principle is a sound guide for human behaviour in that any actions that produce happiness (especially if for the greatest number of people) are good and therefore also ‘moral’. In the hands of a governing authority this principle has an interesting bearing on public decisions and in particular on punishment; whether punishment is necessary in the first place, and if so, what happy consequences it achieves. However, placing happiness as the moral justification of punishment leads to two rather concerning extrapolations: i) the possibility that for the sake of the greater good it may be necessary to punish an innocent person, and ii) punishment being justified only so far as it functions to deter further wrongdoing. The main criticisms levelled at this approach reveal that the good and honourable intentions of utilitarianism are undermined by its misguided correlation between utility and morality. It is chiefly over this point that its moral justification of i) and ii) above is challenged.

Utilitarianism\(^1\) is one approach to what is known as a ‘consequentialist theory’ (Pettit, 2011; Sandel, 2011a), which proposes that the right and moral thing to do depends on the consequences of whatever actions are under consideration. The approach advocates that the end justifies the means if the outcome contributes to the ideal of ‘greatest happiness’. Unfortunately, those ‘means’ do not preclude the possibility that in order to bring the result of greatest happiness some austere or difficult actions may be required; actions that may result in unhappiness or pain for someone in order to benefit the happiness of a greater number of people. It is a theory that sets itself up in direct contrast to non-consequentialist ‘categorical’ or ‘deontological’ morality, that specifies that it is not the case that any means that brings about a happy or greater good result is justified (Pettit, 2011) especially “...if the act violates some other intrinsic moral value” (Sandel, 2011a, 15:22)\(^2\). But the utilitarian sees any categorical response to morality as

---

\(^{1}\) Described here is the branch known as ‘act’ utilitarianism, identified in contrast to the more moderate ‘rule’ utilitarianism. ‘Rule’ utilitarianism holds that certain fundamental rights are sacrosanct and inviolable.

\(^{2}\) This quote is located at 15 minutes and 22 seconds into the presentation.
suffocating of an individual’s ability to respond according to circumstances. By imposing an impractical, absolute and distant concept of morality people remain ‘victims of circumstance’ with no moral mandate to respond according to necessity. A utilitarian embraces the idea that whatever it takes to make the world a better place is justified (Pettit, 2011). To this end he or she cultivates virtues such as honesty, self-control, courage, respect, charity, and generosity (amongst others), because these contribute to happiness. If there arises any situation where these virtues create unhappiness, then a utilitarian would claim that they become immoral and redundant (Pettit, 2011), and dispensable in the face of better personal judgement.

The relevance that his approach has to the issue of punishment is that according to a consequentialist theory there is no point in punishing unless punishing contributes to public welfare and greater happiness. Some of the beneficial consequences of punishing to consider are whether or not the punishment might act as a deterrent by persuading wrongdoers that wrongdoing has unpleasant consequences (Ten, 1987, p. 7), thereby reducing (or hopefully eliminating) crime. If, however, a particular crime has not been prevented, there may be other reasons why punishing is considered. Fines for compensating the victim has the double benefit or restoring (as much as possible) the ‘happiness equilibrium’ for that person, as well as contributing towards any general perception that justice has been served - resulting (hopefully) in a calmer, more peaceful, population. Similarly, public security may be augmented by punishments that remove the wrongdoer from proximity. Ultimately however, a utilitarian would prefer not to punish at all (Tadros, 2011), if the result of non-punishment answers to make the world a better place.

On the face of it there is something appealing about the idea that we should all do whatever it takes to make the world a better place... an attractiveness to an ideal that cares enough to consider the consequences of a reflex as barbaric (Tadros, 2011) as (retributive) punishment, and has the courage to stand up and state that the morality of the matter rests only in the degree of happiness that it produces. But, the problem for utilitarians is this - that however much they might
prefer not to punish in order to reduce suffering in the world, overall not punishing does not reduce suffering; the consequences of not punishing are usually not happy (in the greater scheme of things). The welfare of society dictates the need for punishment (in some form), even if that welfare is only manifest as a deterrent to prevent further crime. Shaved of any other considerations, this approach presents punishment as not about the crime itself (or its consequences), but about the consequences of the punishment. Further extrapolation reveals that to have the desired outcome as a deterrent, the punishment does not need an actual crime or a real criminal. Anyone will do if, by punishment alone, the consequence is that others are deterred.

Critics of a utilitarian approach to punishment base their chief objections around the two concerns just presented. The first criticism points at the theoretical possibility that punishing an innocent person might under certain (albeit extreme) circumstances be acceptable and morally appropriate. The second objection looks at the validity of divorcing punishment from its crime and using it as a deterrent (irrespective of established guilt or innocence) to serve the utility of society. Both these possibilities undermine the benign, gentle and caring face validity of utilitarianism, revealing it as potentially a dehumanising, exploitative and disrespectful approach to morality and justice.

In reply to the first charge, a utilitarian would argue that the pragmatic reality is such that ‘punishing the innocent’ simply does not serve utility (Tadros, 2011). They would cite the manifold ways in which such an eventuality would undermine public confidence in any justice system. This obviously does not make the world a better place, so punishing an innocent person could never be condoned. ‘But what if...’ counters the objector ‘... what if the public does not know that the person is innocent? … what if the public represent a vigilante population baying for blood, and a scapegoat will satisfy that role? What if the innocent has volunteered for the sake of utility?’ To these the utilitarian would patiently explain that the objections are moot; the contingencies for bringing about such extreme outcomes are so remote as to be beyond plausibility (Ten, 1987, pp. 13-14; Tadros, 2011). Such objections are ‘splitting hairs’, but if pressed, yes, it would be the moral thing to do.
In reality though, the utilitarian’s claim that such an eventuality, although morally appropriate in theory, would never actually happen, is spurious. There are countries where (for as shallow yet powerful a motivation as ‘saving face’ for example) an innocent person might be coerced into confessing and ‘sacrificed’; countries that are constantly under international scrutiny and censure for their violations of ‘human rights’ principles. These countries openly justify their position, claiming that they act for the benefit of the greater public good, to protect confidence in the judicial system, and deter crime. In countries where such an approach is applied, violation of human rights such as punishing the innocent will continue while it is considered morally appropriate. Yet even if it were not... even if the possibility were in fact pragmatically impossible, no moral or ethical approach can ever justify, even if only in theory, humans behaving in a way that is so disrespectful to others (Pettit, 2011, citing Kant) or mandate a political authority knowingly allowing such violations.

This criticism is echoed in the second charge, that of justifying a punishment solely as a means to deter. The purpose of punishing as a deterrent is twofold; to discourage the wrongdoer from repeat offending, and to provide a general deterrent within society (Ten, 1987, p. 7). The criticism of using punishment as deterrence is based on the premise that human dignity is an inviolable right; and that for punishment to be effective in this way would mean that it would need to either very painful, or very humiliating, or very depriving. Such strategies for deterrence are not appropriate as a means to an end, and not particularly conducive to making the world a happier place. The utilitarian would immediately counter that indeed if such a measure did not make the world a happier place, then it would not be condoned (Pettit, 2007), but that a criminal (if truly guilty) has forfeit his or her right to being treated with the usual expected human dignity, … in fact nobody has the ‘right’ to be treated with dignity anyway; human dignity is only a natural and lucky by-product of a utilitarian approach. Choosing to treat people and their property with dignity and respect are ‘goods’ that make the world a happier place. Likewise, if not doing so is required to protect the happiness of society, it is morally justified, and using punishment as a means to deter is a practical example of that principle at work.
This rebuttal is a little more difficult to challenge. The nature of crime and the need to prevent it is perennially relevant, and draws an instinctive response that guilty people deserve to be punished, and if this deters them or others from the same, all the better. It makes sense, but ‘begs the question’ by presupposing the strategy might work in the first place. Ten (1987, pp 8 - 10) cites research that disputes the effectiveness of deterrence on the wrongdoer, and questions its influence in general. Dupré (2007) and Pettit (2011) both challenge the idea as ineffective, particularly for certain types of crimes. What perpetrators fear is being caught, rather than being punished (Dupré, 2007, p. 195). Of course, if it does not work the utilitarian would say it should not and would not be considered. But, this is not the point. The real issue is not that people are deterred from crime, but that deterrence as the only purpose of punishing cannot be considered ‘moral’. Deterrence may be a by-product of punishment, a collateral benefit if you like, but punishment is about justice, and deterrence is not a function of justice.

Beginning to emerge within this discussion are concerns centred on the underlying assumption that ‘utility’ equals morality or justice, and represents the ultimate ‘good’. Philosophers such as Kant (cited by Sandel, 2011b) and Williams (cited by Pettit, 2011) would argue that accepting a moral code that is contingent on circumstances is to admit moral defeat, and that individuals taught to accept such a maxim have lost their intellectual freedom and self-determination. They become mere “object[s] of experience” (Sandel, 2011b 5:39), in which moral decisions are ‘heteronomous’ (Kant, 1957, pp. 336-337) to whatever external contingency is delivered to the mind via the senses (Sandel, 2011b). A utilitarian theory creates a precept of ‘moral opportunism’ (Pettit, 2011) with no value placed on moral integrity. The abhorrent and repulsive become justified once the desired outcome is reached (Pettit, 2011, citing Williams). Dupré (2007, p. 66) explains the way in which such an ideology degenerates. He says:

Problems can arise … when an inappropriate end is chosen and the choice is made in light of ideology or dogma. If a political ideologue, … sets up a particular end as important to the exclusion of all others, it is a short step … to conclude that it is morally acceptable to use any means whatsoever to achieve that end.
As a foundational concept for justice and morality, utility’ is an inappropriate end, and potentially becomes an excuse for callous expediency and even the basis of systemic corruption. The ‘end’ may be perceived as utility, but it is not moral. Kant (1957, p. 268) identifies morality as beyond the nature of the empirical world, existing rather as an end in itself, governed by one’s ability to engage in metaphysical habits of “pure reason” (Kant, 1957, p. 269) making morality a universal and transcendent reality to which even God is subject, and which alone can enable self-determined and autonomous moral decision-making.

The consequentialist desire to be responsive to circumstances is so equivocal as to make the theory self-defeating. The argument looks like this: utility is not served by punishing the innocent, unless utility is served by so doing. Likewise, neither is utility served by not punishing the guilty, unless utility is served by so doing. A utilitarian theory of moral justice cycles through a decaying spiral to self-annulment. Its eventual message is that the most appropriate form of utilitarian justice is not to be utilitarian at all, because that in the end does not serve utility. By inviting discourse on how this world can be made a better place consequentialist theorists have done a good job, and are nearly convincing. There is appeal in the idea of formalising morality to serve the happiest possible outcome. But a stronger bottom line than utility is needed for establishing a more structured and less corruptible moral foundation, especially in response to the problem of crime. Society benefits from knowing that under no circumstances is punishing the innocent ever going to be justified as morally appropriate. Likewise, punishing solely for the purpose of deterring misses the point about the nature of justice and the role of punishment.
References:


