THE MORAL MURDERER. A (MORE) EFFECTIVE COUNTEREXAMPLE TO CONSEQUENTIALISM

Eduardo Rivera-López

Abstract

My aim in this paper is to provide an effective counterexample to consequentialism. I assume that traditional counterexamples, such as Transplant (A doctor should kill one person and transplant her organs to five terminal patients, thereby saving their lives) and Judge (A judge should sentence to death an innocent person if he knows that an outraged mob will otherwise kill many innocent persons), are not effective, for two reasons: first, they make unrealistic assumptions and, second, they do not pass the rule-consequentialist institutional test. My example (The Moral Murderer), instead, assumes a realistic empirical framework and the relevant action does not undermine basic social institutions. On the contrary, it reinforces them. In The Moral Murderer, Tom (an adult male) is morally allowed to murder a person (preferably a woman) in order to be punished to death.

I

Among the many reasons consequentialism has been criticized, one is that it may allow (or require) us to perform morally atrocious acts. Here are two well-known examples:

> Transplant: A doctor is morally allowed to kill one person and transplant her organs to five terminal patients, thereby saving their lives.[1]

> Judge: A judge is morally allowed to sentence to death an innocent person if he knows that an outraged mob will otherwise kill many innocent persons.[2]

---

[1] See Thomson (1976), p. 206; also Thomson (1990), p. 135.
[2] See McCloskey (1957), pp. 468–469; Foot (1967), p. 270; Nozick (1974), p. 28. In my formulation of both examples, the agent is morally allowed but not necessarily required to perform the action. This is enough to show the counter-intuitive character of these examples. More on this qualification in note 7.

The argument is simple: Consequentialists (at least those who value several lives more than one) should accept Transplant and Judge. Since the acts involved in both cases (killing in Transplant, sentencing in Judge) are intuitively morally atrocious, consequentialism is flawed. As we will see, there are at least two common answers to this objection. The first one (endorsed by act-consequentialists) is to claim that the occasions in which an agent (a doctor or a judge) would be allowed to do such horrible things are so extremely exceptional that the examples turn out to be unrealistic (in a sense to be defined below) and, therefore, not effective. The second answer is to reject that reasonable versions of consequentialism really allow such terrible actions in any circumstance (exceptional or not). This answer involves the endorsement of some kind of *indirect* consequentialism, for example, rule-consequentialism.[3]

I will not object to these strategies. In fact, I will assume that they are successful and that, therefore, examples such as Transplant and Judge do not threaten consequentialism. For a counterexample to consequentialism to be effective, it should not depict an exceptional or unrealistic scenario; nor should it be such that the relevant action performed in the example (killing, or sentencing, for example) goes against some optimizing rule. My purpose in this paper is to provide a counterexample that is effective in both senses.

II

Before offering the counterexample, it is worth giving a closer look to the arguments of act- and rule-consequentialists, according to which Transplant and Judge are not effective counterexamples.

---

[3]   There is a third answer: abandoning agent neutrality. This implies believing that consequentialism and agent-relativity are compatible, and that agent-relative consequentialism is plausible (according to this view, we are morally required to bring about the best consequences from our relative individual position, not from the point of view of a neutral evaluator). Some authors have defended the compatibility (see Broome (1991), p. 8 and 13), and a few the plausibility of this view (see Portmore (2001), (2003), and (2005)). For a critical assessment of this project, see Schroeder (2006) and (2007). Agent-relative consequentialism goes against the traditional account of consequentialism, according to which the right act is the one that maximizes value from an objective, neutral point of view. I will not discuss agent-relative consequentialism in this paper and I must admit that my example is not effective against it.

Act-consequentialists think that Transplant, Judge – and other scenarios of the same kind – are thought experiments in which so many unrealistic features are incorporated that virtually no actual situation satisfying such features could occur. The examples are unrealistic in the sense that, in these scenarios, the way in which social practices and institutions work in the real world is ignored. In the real world, if doctors kill healthy bystanders or other patients and transplant their organs, the purpose of those practices and institutions (hospitals) is undercut and consequences, in the long run, can be worse, for example, because people will lose any confidence on doctors and hospitals. If judges condemn innocent persons, the same is likely to happen: confidence in the judicial system will be destroyed and the deterrent effect of punishment seriously undermined. Again, consequences, in the long run, will be worse. It may be true that under unrealistic circumstances, doctors should transplant and judges sentence. Unrealistic circumstances are possible worlds in which foreseeable causal consequences of our actions are different from those we normally associate in our real world. The assumption of these unrealistic circumstances makes these counterexamples fail, because, when we imagine Transplant or Judge in order to offer an intuitive reaction to the relevant action (the transplant and the punishment to the innocent), we cannot avoid such a reaction to be influenced by our everyday assumptions about how institutions and social practices work and about the risks of breaking the rules of those institutions.[4]

The rule-consequentialist's response is different. The problem with scenarios such as Transplant or Judge is not their artificiality. The problem is, rather, that rules that allowed the relevant actions, if everyone accepted them, would have bad consequences in the long run. A rule that allows judges to condemn innocent persons or doctors to kill healthy bystanders, if generally accepted, would not have optimific consequences.[5]

Notice that sophisticated act-consequentialism and rule-consequentialism share something important. It is what I will call

---

[4]   On condemning an innocent, see Sprigge (1965), pp. 275–276 and Smart (1973), pp. 69–73. On the transplant case, see Griffin (1996), p. 99. Griffin's argument is more epistemic. The doctor cannot know the consequences of his departure from the rule of the practice: 'they [doctors] cannot know whether, or how widely, people will learn what they have done. They cannot know how much fear and mistrust that knowledge will produce'. (p. 99).

[5]   For a well-known version of this argument, see Rawls (1955), pp. 25–29.

'the institutional argument'. In both approaches we find the idea that performing certain actions will go against a beneficial set of rules, institutions or practices. The only difference between the responses is that the nature or role of these institutions, rules or practices is different. For act-consequentialists, they are simply rules of thumb or conventions that are useful in promoting the best consequences. For rule-consequentialist, they are more than that: they are genuine moral rules, which are supported on the basis of the indirect consequentialist's argument that general acceptance of these rules produces the best consequences.[6]

In any case, the main problem of Transplant and Judge is that the relevant actions go against well established and beneficial institutions or practices (doctors in a hospital, judges in a court). This is problematic, either because it makes the examples unrealistic in the mentioned sense (they assume that actions do not have the effects they normally have within those contexts) or because they go against institutions that are, in themselves, consequentialistically optimific. A more successful counterexample should be one that, first, assumes realistic causal connections and institutions working as they do in the real world, and, second, assumes institutions that are defensible from the consequentialist point of view.

III

The example I want to discuss is inspired by the idea, shared by many experts and lay people, that capital punishment has an important deterrent effect. As I will soon show, some serious studies claim that each execution deters a number of persons from killing. In that sense, executing a capital murderer saves the life of several innocent persons. This is why consequentialists often support death penalty. With this idea in mind, consider

The Moral Murderer:
Tom is an adult male. He lives in a state of the U. S. where capital punishment is routinely enforced (say, Texas or

---

[6] Another way to explain the convergence on the institutional argument is that both sophisticated act-consequentialism and rule-consequentialism share the rule-consequentialist decision procedure, according to which agents should decide what to do by applying rules. Both kinds of consequentialism differ on the criterion of wrongness (see Hooker (2008)).

Oklahoma). There he kills an innocent, white, person (preferably a woman), involving a number of aggravating circumstances. Tom does not make much effort to escape justice. After committing the murder, he does not show remorse. He is poor and, therefore, cannot afford a private lawyer. Public defenders are not especially good. Since his probable execution will deter other murderers and thereby save the lives of other innocent persons, Tom is morally (at least) allowed to kill.[7]

My claim is that The Moral Murderer is an effective counterexample to both act- and rule-consequentialism. To qualify as such, the relevant action described in the example should be realistic and should presuppose only plausible empirical claims. The main claim is the one I suggested above: that capital punishment has a positive deterrent effect. More precisely, I will assume that each additional execution deters at least 18 murders (more than life imprisonment). I will call this claim 'Main Empirical Claim'. Since the Main Empirical Claim will surely strike some readers as a strange or controversial empirical assumption, a brief explanation is required.

First, the Main Empirical Claim is consistent with a substantial part of the most recent empirical literature on capital punishment. Econometric studies evaluating the deterrent effect of capital punishment vary quite strongly, and it is not my purpose here to review all the relevant literature. Nevertheless, it is worth mentioning that one of the most serious empirical studies on the issue claims that 'our most conservative estimate is that the execution of each offender seems to save, on average, the lives of eighteen potential victims'.[8] Second, and more important, I am *not* claiming that the Main Empirical Claim is true. In fact, there are many serious statistical studies claiming that it is not.[9] However, for my purposes it suffices that it is plausible: neither

[7]   As in the original examples, Transplant and Judge, I weaken my claim somewhat by saying that Tom is morally (at least) allowed (but not necessarily required) to kill one innocent person. This enables me to avoid dealing with the problem of overdemandingness and to include into my target also some hybrid moral theories that accept agent-relative options, such as Samuel Scheffler's theory (Scheffler (1982)).

[8]   Dezhbakhsh et al. (2003), p. 373. Note that this study compares death penalty with the alternative punishment applied during the death penalty moratorium in USA (1972–1976), life imprisonment.

[9]   See for example Donohue and Wolfers (2006).

ridiculous nor far-fetched. Given that it is supported by at least some respectable work, we should take it that it is, at least, plausible.[10]

As in the case of the Main Empirical Claim, other key empirical claims in this paper will also be supported by serious empirical literature or by available statistics. Once again, I do not pretend to truth or accuracy. All I assert is plausibility. Moreover, when empirical research or statistics do not provide what I need (or better: when I do not find in such literature what I need), I speculate. This is not problematic, since, I repeat, all my argument needs is that the empirical background of my example is consistent with serious social research and does not appeal to extraordinary, unreal, or far-fetched circumstances or causal connections (as Transplant and Judge do, according to sophisticated act- or rule-consequentialists).

IV

Taking the Main Empirical Claim for granted, we should now consider some facts about murder and capital punishment in the U.S. Even if the Main Empirical Claim were true, the plausibility of The Moral Murderer still depends on the probability that a person who has committed a capital murder will ultimately be executed. The probability is quite low. Between 1977 and 2003, roughly 120,000 capital-eligible murders were committed.[11] In the same period, 885 persons were executed.[12] That means that the relevant probability of being executed has been 0.74%. The probability is higher if we focus on Texas. There were about 10,000 capital crimes in Texas between 1977 and 2003 and 313 execu-

---

[10] By being the Main Empirical Claim 'plausible' I do not mean it to be 'likely' true. I mean that reasonable people, on the light of some evidence, may believe it to be likely true. This is, in my view, quite close to say that the Main Empirical Claim is not ridiculous or farfetched. For example, Sunstein and Vermeule, in an important (and controversial) paper on capital punishment, take Dezhbakhsh's study as true (Sunstein et al. (2005–2006)). This supports the idea that such study is plausible (in my sense), not that the study is likely true.

[11] Fagan et al. (2006), p. 1819. Fagan estimates that about 25% of homicides are death-eligible. This is consistent with the estimate that between 10% and 15% of the persons arrested for homicides have committed death-eligible crimes (see Baldus et al. (1997–1998), p. 1656).

[12] See http://www.deathpenaltyinfo.org/executions.

tions.[13] The relevant probability has therefore been 3.13% in Texas. The state with the highest execution rate is Oklahoma, where the relevant probability has been nearly 5%.[14] These statistics might seem to show that The Moral Murderer is a non-starter: Tom's crime is sure to cost one person's life, and has only a very low chance of saving eighteen persons' lives (less than one percent in the country at large, 3.3 percent in Texas, 5 percent in Oklahoma).

Before analyzing how further empirical facts might help to my argument, it will be useful to briefly review the way in which consequentialist theories approach the problem of making moral decisions under risk. In order to exclude irrelevant considerations, let us focus on the case of *saving* persons from death. Imagine that you have to choose between saving one person's life (A) or two persons' lives (B and C).[15] A consequentialist would surely say that you should save B and C. Now imagine that you have to choose between saving A's life and having 50% chance of saving B's and C's lives. If we assume what we may call 'moral risk neutrality', you should be indifferent. Now suppose that the chance of saving B's and C's lives is not 50% but 60%. Would a consequentialist say that you should try to save B and C (with a 60% chance), although you can save A for sure? I am not sure, but I do not know any (consequentialist) argument for abandoning moral risk neutrality and endorsing a degree of moral risk aversion.[16] If our fundamental moral command is to optimize consequences (which, in this case, means maximizing the number of lives saved), it would be strange to claim that we should follow a course of action that does not maximize the *expected* number of lives saved.

Assuming moral risk neutrality, we can say that, according to consequentialism, we are morally required to save B's and C's lives (and thus to let A die) whenever the chance of saving B's and C's lives is more than 50%. By the same token, we can also say that a greater than 25% chance of saving four persons, or a greater than

---

[13]   Fagan et al. (2006), p. 1821; see also http://www.deathpenaltyinfo.org/executions.
[14]   There have been about 7000 homicides between 1976 and 2005 in Oklahoma (*Bureau of Justice Statistics*). If, as assumed, 25% of them are capital murders, we have about 1750. In the same period, there have been 83 executions: 4.75%.
[15]   Other things are equal in all these examples.
[16]   Parfit describes consequentialism explicitly as a risk neutral theory (Parfit (1984), p. 25). Broome discusses the point and notes the problems of risk neutrality in moral decisions. However, he does not abandon it (see Broome (1991), pp. 124–126).

12.5% chance of saving eight persons, or a greater than 6.25% chance of saving sixteen persons, are each morally preferable to saving one person's life for sure.

Although I do not know how a consequentialist could reject moral risk neutrality and endorse some degree of risk aversion, we should notice that risk neutrality can lead to very unpalatable conclusions. For example, it implies that we should opt for a course of action that has only a one percent chance of saving 101 persons, instead of saving one person for sure. It is not my purpose to defend any view on this matter. I am only trying to offer a plausible account of consequentialism, even if it is not particularly favourable to my purpose (which is to show that The Moral Murderer is an effective counterexample to consequentialism). I will therefore assume that some degree of risk aversion is defensible on consequentialist grounds. Although I do not have any way of determining what particular degree would be acceptable, I think the following should be a sufficient concession. We have seen that, under risk neutrality, you should choose to try to save 16 persons and let one person die if the probability of saving the 16 persons is greater than 6.25% (and your chances of saving the one person are 100%). This entails that if the chances of saving the 16 persons is, say, 6.3%, you should proceed. Contrary to risk neutrality, I will assume that, in order to be allowed to choose to try to save the 16 persons, the chance should be 13% (slightly more than twice the 'indifference percentage'). This percentage is certainly somewhat arbitrary. However, I do not think this should be regarded as a problem, since it should be interpreted as a concession: I do not find any reason to abandon risk neutrality at all and, as it will be obvious, any departure from risk neutrality in favour of risk aversion will tend to *undermine* the effectiveness of The Moral Murderer as a counterexample to consequentialism.

Returning to the case of capital punishment, and applying the above criterion, I will assume that the consequentialist will claim that a necessary condition for The Moral Murderer to be a counterexample is that the relevant probability of execution of a capital murderer is at least 13%. Recall that, according to the Main Empirical Claim, one execution saves 18 lives. If Tom kills someone, thereby incurring a 13% chance of being executed, he will be sacrificing one life for sure and another life with a 13% chance (his own life), in order to have a 13% chance of saving 18 persons. The cost of proceeding in The Moral Murderer would be

1.13 (1 + 0.13) expected lives, whereas the gain would be 2.34 (18 × 0.13) expected lives.

Now, as we have seen, the probability of being executed for a capital crime in states such as Texas or Oklahoma oscillates between three and five percent. It seems that even in these states we are still far from the (self-assumed as) required 13%. However, there are two crucial additional observations to make. First, the execution rate is an average estimation that takes into account *all* capital murders. But Tom could easily raise his probability of being executed above the average in several ways. Some of them are suggested in the example itself. For example, there is empirical evidence that, other things being equal, the probability of being sentenced to death is higher if the victim is white.[17] It is also quite clear that by performing a truly heinous murder, with many aggravating conditions, and in the course of committing another crime (rape or torture, for example), Tom's chances of being sentenced to death and ultimately executed can be raised still higher. For example, there is evidence that, in the case of rape, the chances of being prosecuted and condemned can be increased by choosing a victim that does not engage in 'risk taking behaviour'.[18]

Although there is empirical evidence that all these factors increase Tom's chances of being executed, I do not have any scientifically grounded empirical estimates as to *how much* the chances increase. I can only speculate in this regard. However, in view of the partial information available on these factors, I think that, even on an extremely conservative estimate, the general 3% to 5% chance of being executed for a capital crime could well increase two or three times if the additional conditions are met. At least it is not unreasonable to assume so.

A second point is still more relevant. The current execution rate is very low, but it might well be much higher. Consequential-ist defenders of capital punishment should be happy to increase

---

[17]   According to GAO (1990), '[I]n 82 percent of the studies, race of victim was found to influence the likelihood of being charged with capital murder or receiving the death penalty, i.e., those who murdered withes were found to be more likely to be sentenced to death than those who murdered blacks'. (p. 5). For one of these studies, see Baldus et al. (2003–2004), pp. 1423–1426, and Baumer et al. (2000), p. 304. The same studies show that the race of the defendant was also a relevant factor in the pre-*Furman* era, but there is no evidence of systematic discrimination against black defendants after 1976.
[18]   In the case of rape, the chances of being prosecuted and condemned are higher if the victim did not engage in 'risk taking behaviour' (see Spohn et al. (1996), p. 677).

the execution rate. In fact, empirical evidence suggests that the deterrent effect of capital punishment increases when the ratio of executions to death sentences increases.[19] We can therefore imagine a realistic scenario in which death sentences are more systematically brought to a 'successful' conclusion in actual executions. Imagine that 80% of all those who are sentenced to death are in fact executed (which is by no way unrealistic) and take again the case of Texas, where 819 persons were sentenced to death between 1977 and 2003.[20] As we have already seen, 313 persons were executed during the same period (38%). Let us then imagine that, instead of 38%, 80% of those sentenced to death were executed. That would have resulted in 655 executions. Since there were roughly 10,000 capital murders committed over the same period, the execution rate would have been 6.55% (rather than 3.3%). The same exercise is possible at the national level. Between 1977 and 2007 there were 7348 death sentences and roughly 120,000 capital murders. If 80% of those sentenced persons had been executed, the execution ratio would have been 4.9% (rather than 0.78%).[21]

This is not all. Increasing the death sentence/execution ratio is not the only way to increase the capital murder/execution ratio. We might realistically imagine a system in which a higher number of persons who are prosecuted for capital murder are sentenced to death. Again, increasing such proportion should increase the number of capital murderers who are eventually executed.

All these considerations should be sufficient to justify the belief that, even on a conservative estimate, Tom's probability of being executed in The Moral Murderer might well be much higher than 13%, either in the real world or in a realistically imaginable world (one that would also be *desirable* for consequentialists). Moreover, recall that the 13% requirement is the result of granting, for the sake of argument, a certain degree of moral risk aversion, which is not obviously supportable on consequentialist grounds.

---

[19]   For example, one study claims to show that 'one fewer murder is committed for every 2.75-year reduction in the expected death row wait' (Shepherd (2004), pp. 314–315). If this is true, it seems clear that additional executions of death sentenced persons would increase the deterrent effect.

[20]   http://www.deathpenaltyinfo.org/article.php?scid=9&did=847.

[21]   http://www.deathpenaltyinfo.org/article.php?scid=9&did=873.

V

In order to argue that The Moral Murderer is an effective counterexample, I am not only assuming that all the empirical claims I have relied upon are at least reasonable, but also that a consequentialist (who also believes that all these claims are plausible) supports the morality of capital punishment. The real world is one in which some people break moral rules by (for example) killing other people, so that the society must have institutions to deal with non-compliant persons: for example, by imposing capital punishment.

Bearing this in mind, I can finally argue that The Moral Murderer is a more effective counterexample to consequentialism than Transplant and Judge. Firstly, it rests on plausible empirical premises (the Main Empirical Claim and others). This is crucially different from Transplant and Judge. In these more standard examples, the relevant actions (killing, transplanting organs, and sentencing innocents) are anomalous. They are actions that are not normally performed by the relevant kinds of agents (doctors or judges) in the actual world. In contrast to this, in The Moral Murderer, Tom is a simple man who, like many others, can commit a murder and be subject to criminal prosecution. There is nothing unrealistic here. More than 4,000 murders are committed each year in the U.S. that are eligible for the death penalty.

Secondly, Tom's behaviour in The Moral Murderer does not undermine any social rule, institution or practice. Therefore, the institutional argument does not apply. In the real, non-ideal world, thousand of persons kill other persons every year. Many of them are prosecuted and a few of them are executed. This is the institutional framework assumed by the example. In the real, non-ideal world, the consequentialist supports this framework.[22] Tom's action in The Moral Murderer does not undermine (on the contrary: it reinforces) capital punishment, which is the most relevant institution involved.

These two points should suffice to show that The Moral Murderer represents a clear improvement as counterexample against

[22]   The Moral Murderer has a further advantage: it does not require the creation of a new institution, which can collide with existing practices or norms and, therefore, frustrate the aim of maximizing good consequences. This is one of the problems of John Harris's 'Survival Lottery' (see Harris (1975); for consequentialist objections, see Singer (1977)).

consequentialism. To be sure, many objections can be made. Considering some of them will help to clarify and (hopefully) strengthen the argument.

Some objections appeal to the (bad) consequences of The Moral Murderer. If the moral murder is a morally permissible act for Tom, it also is permissible for every moral agent. We might think that, in a world plenty of moral murderers, we would be afraid of being killed by one of them. Consequences, in terms of aggregate wellbeing, might well be worse. This is similar to the consequentialist objection to Transplant that our fear to be killed by a doctor would make medicine impossible. However, we must stress that, if the factual argument is plausible, the probability to be killed is lower in a world with moral murderers than without them, since moral murders prevent (some) people from being killed. Note that the number of moral murderers will surely be very low, given the high personal cost of being one. The fear of being a victim of doctors and judges in Transplant and Judge has to do with the fact that doctors and judges are not supposed to kill or sentence innocents. Instead, murderers are supposed to kill. If the existence of some (few) moral murderers diminishes the number of murders, we have no reason to be afraid. To the contrary, we should feel safer.

A second concern with the moral permission to perform moral murders is that it could make the moral rule against killing weaker by incorporating a new exception to that rule. However, we already have some exceptions to the prohibition to kill. We are allowed to kill in self-defence or in defence of others. Do these exceptions weaken the prohibition? Maybe. Note, however, that, for the same reason I mentioned above, the weakening effect on the prohibition should be lower in the case of the moral murder than in the case of self-defence: almost everyone is likely to defend herself from an attacker, but only a few are likely to commit a moral murder.[23]

Another problem is that allowing moral murders would weaken the self-proclaimed monopoly of coercion by the state. It would be a kind of permission to take justice into one's own hand. The bad consequences of renouncing to such monopoly are obvious. But The Moral Murderer does not imply such a renunciation. Legally,

---

[23]   In this respect, the case of the moral murderer is more similar to the case of killing someone in defence of a third person and thereby risking one's own life, which is considered permissible both by law and morality.

the action performed by Tom is prohibited and must be punished. The example does not pretend to show that the consequentialist must agree with the legal permission of Tom's action but only with the moral one.

A further concern is that, in a society with moral murderers, 'real' potential murderers would not be deterred by capital punishment, or at least not as much, since they would be aware that some of the executed persons are sacrificing themselves *in order to* deter real murderers. I think this is not a real problem. Statistics are opaque about why people become deterred by death penalty. The only important point seems to be the perception that, if I kill, I can be executed, on the basis of observing that other persons that kill are executed. The deterring mechanism does not require potential murderers to know why previous executed murderers were executed.

Further problems are related to Tom's psychological features. He is a kind of moral fanatic. If the judge discovers this feature, Tom would be declared insane and, therefore, not executed. If we assume that consequentialism is right, I do not see any reason to consider Tom as a moral fanatic, rather than as a moral hero. A consequentialist judge should therefore execute him, as a way of allowing him to accomplish his moral action. Moreover, Tom would obviously not tell the judges that he is moral murderer.

A related point is that moral permission to commit a moral murder might induce to moral fanaticism (or 'heroism'), which would in itself carry out bad social consequences. One might think that Tom's action could encourage others to do the same, which, at the end, might have catastrophic consequences. However, there is no straightforward connection between permitting an action and inducing such an action. For example, we allow people to perform heroic actions (for example, sacrificing one own life to save others in war). However, this does not seem to induce many people to perform those actions. In fact, heroic actions remain rather exceptional. There is no reason to think otherwise in the case of moral murderers, which amounts to an enormous personal cost.

One might also object that my argument makes a misleading use of statistics. In order to support The Moral Murderer, those statistics should not only show that, *on average*, some number of innocent persons for each execution are saved, but that each *additional* execution will save an additional number of people. And they cannot provide that much. However, even if it is true

that statistics on death penalty can only show the average number of persons saved, this is enough for a consequentialist. For her, a kind of action that saves five lives for sure would be as required as an action that saves an average of five lives (sometimes that kind of action saves more, sometimes less).

There is a difference between The Moral Murderer and Judge or Transplant that might suggest a further objection. In Judge and Transplant the action is supposed to be the *only* way of saving lives. The case of Tom in The Moral Murderer seems to be different. Given the length of time prisoners are on death row, and the cost of the process, his action is an extremely inefficient and uncertain way of saving lives. He could save far more lives over that period by, for example, working and giving part of his wages to provide poor countries with clean water and food. So consequentialism would say Tom's action in The Moral Murderer is not optimific and therefore is wrong.[24]

This objection has two parts. The first one concerns the cost of the process, which could be saved and spent to rescue other (perhaps thousand) people from death (for example, from starvation). The objection is strong only provided that the money would be spent in assistance, had Tom not committed the moral murder. However, I think it is realistic to think that this is not the case. Moreover, this is completely beyond Tom's control. The second aspect concerns the money that Tom could himself spend in saving other lives. I have tried to depict a picture that is, at the same time, realistic, and consistent with Tom's being unable to save persons in other ways: he is poor and, without committing the moral murder, his foreseeable economic position will at most be sufficient to cover his own living expenses. Still, let us suppose that this is not realistic: Tom *could* work hard and spend some money to save much more people than by committing a moral murder. I think the example does not become less effective, for the following reason. By committing a moral murder, Tom certainly makes it impossible to save starving people in other ways. But millions of persons do not save people in either way. They are affluent enough to donate money to Oxfam or otherwise rescue people from starvation, but they do not. They are, according to the consequentialist, blameworthy. The Moral Murderer forces the consequentialist to claim that Tom is *less* blameworthy than those

---

[24]   I owe this objection to an anonymous referee.

millions persons (at least he saves eighteen persons). This is enough, in my view, to make The Moral Murderer successful as a counterexample to consequentialism, since we intuitively think that he is not less, but *much more* blameworthy.

Let me finally consider some objections made from a rule- or indirect consequentialist outlook. The first one is rather obvious: Tom's killing in The Moral Murderer defeats a fundamental moral rule: the rule 'don't kill'. Such a rule, if universally accepted, has the best consequences. Therefore, The Moral Murderer would not be acceptable to a rule-consequentialist. One answer to this argument is a variation of the traditional objection to agent-relative constraints against killing: Tom's killing in The Moral Murderer does not defeat the rule against killing, because it contributes to less instances of killing (it saves more persons from being killed).[25] Tom's killing reinforces the rule. This argument may well be ineffective against deontologism, because agent-relative reasons are immune to any kind of aggregation. But it seems quite strong against rule-consequentialism, because the main concern for rule consequentialists is the degree of observance of the optimizing rule. There is an additional, crucial, point. Tom's action in The Moral Murderer is performed in a realistic, non-ideal, scenario. That means that the rule against killing is in fact being broken by many people. Capital punishment is one measure, among many others, for dealing with this unfortunate fact. If the rule-consequentialist thinks that moral rules should not change *at all* in the real, non-ideal world, she should also reject capital punishment, which also breaks the rule against killing (it *consists* of killing). More plausible for a rule-consequentialist is to accept some accommodation of moral rules to the fact of partial compliance. One such accommodation is accepting death penalty as an institution oriented to diminish the number of murders. If this is an acceptable accommodation, I see no reason to reject the morality of an action that reinforces such institution by helping to diminish the number of killings still more.

One might insist with a more sophisticate view of rule-consequentialism. According to this view, morality consists of a complex set of rules, which also specify conditions that justify their violation. Rules against killing include clauses providing

---

[25] This is the argument that many have used against Williams famous example of the person that credibly threatens to kill twenty persons if you do not kill one person (Williams (1973), pp. 98–99). See, for example, Scheffler (1982), pp. 84–90.

exceptions in partial-compliance situations. Given our knowledge of human nature and the tendency to violate rules, a maximizing set of rules would be likely to allow the state to violate certain moral rules (for example, against killing) that they would not permit individuals to violate. This explains why the state is allowed to punish murderers with death but Tom is not allowed to commit a moral murder.[26]

In order to make this objection plausible, we should provide an independent explanation as to why the human tendency to violate rules gives us a reason to prohibit Tom to perform an optimizing killing (committing a moral murder), and, at the same time, it does *not* give us such a reason when the state does the same (executing a murderer).[27] Such an explanation is not easy to provide since Tom's action in The Moral Murderer (contrary to the relevant actions in Transplant and Judge) is a simple action, which is itself a violation of a rule. Note that killing is already the worst thing one can do. Therefore, we should not fear that Tom could overpass the limit of the permissible action (killing someone) and do something wrong (in the sense of having bad consequences).[28]

My aim in this paper has been to develop an effective counter-example of act- and rule consequentialism conceived as agent-neutral theories.[29] One might wonder whether my example works against other versions of indirect consequentialism, such as virtue-consequentialism. According to this theory we have the moral duty to cultivate right desires or a virtuous character, because this is the best way to optimize consequences. Fulfilling such a duty might be thought to be incompatible with performing the horrendous action of The Moral Murderer. This might be true. If so,

[26] I owe this objection to Joshua Gert.

[27] There is a difference between both killings: moral murderers kill innocent persons whereas the state kills (or executes) only guilty ones. However, such a difference is irrelevant for consequentialism. The consequentialist justification of punishment is based on deterrence, not on retribution to the guilty. Moreover, we usually consider that killing is wrong, regardless of whether the victim is innocent or not. The only exception seems to be self-defence. In that case, however, most people think that we are also allowed to kill an *innocent* attacker.

[28] Imagine that Tom's action promotes other moral murders committed outside the jurisdiction where the execution ratio is 13% or more. This would bring about bad consequences. However, we can again imagine a world in which the required ratio holds in the entire country. Although this move turns the example hypothetical, it does not undermine it, because the hypothetical world is not unrealistic (as it is a world in which doctors kill bystanders, safely transplant their organs and no further negative consequences follow). Moreover, such a world would be welcomed by the consequentialist.

[29] In footnote 3 I set aside agent-relative consequentialism.

my argument would not apply to this kind of moral theory. However, I would suggest that the following argument, similar to the one just presented against the rule-consequentialist, applies here as well. In a non-ideal world, it is not clear that the virtuous character is one that prohibits us to kill in *any* circumstance. If this were the case, the death penalty would also be impermissible. If we accept death penalty (as I assume the consequentialist will), we do not have any reason to reject the idea that a virtuous person could kill in The Moral Murderer. Recall that, in The Moral Murderer, Tom is ready to sacrifice his own life to save others. If this is Tom's motivation, I do not see any reason to think that he is not a virtuous person!

Conclusion

What does my argument ultimately show? If plausible, it improves on some arguments against consequentialism that are based on our moral intuitions about the impermissibility of killing, torturing, or otherwise violating individual rights. A number of these arguments are supported by examples like Transplant and Judge. Consequentialists of different kinds have objected to these examples, claiming that they are unrealistic or that they do not take into account institutional consequences. If The Moral Murderer is plausible, it undermines this kind of objection: it is not unrealistic and does not have negative institutional consequences. Of course, this is only part of the discussion. A consequentialist might have a different kind of reaction to Transplant and Judge (and, a fortiori, to The Moral Murderer): she might simply reject our moral intuitions, and claim that the doctor is allowed to perform the transplant and that the judge is allowed to sentence the innocent. If my argument is sound, this kind of consequentialist should regard it as morally permissible for Tom to kill an innocent in The Moral Murderer. In that case, the example will have the virtue of helping us to focus the discussion on the essential clash between irreducible consequentialist and deontological moral intuitions.[30]

Law School, Universidad Torcuato Di Tella/CONICET
Miñones 2159
1428 Ciudad Autónoma de Buenos Aires
Argentina
erivera@utdt.edu

References

Baldus, D. et al. (1997–1998). 'Racial Discrimination and Death Penalty in the Post-*Furman* Era: An Empirical and Legal Overview, with Recent Findings from Philadelphia', *Cornell Law Review*, Vol. 83, pp. 1638–1770.

—— (2003–2004). 'Race discrimination and the Legitimacy of Capital Punishment: Reflection on the Interaction of Fact and Perception', *DePaul Law Review*, Vol. 53, pp. 1411–1496.

Baumer, E. P. et al. (2000). 'The Role of Victim's Characteristics in the Disposition of Murder Cases', *Justice Quarterly*, Vol. 17, No 2, pp. 281–307.

Broome, J. (1991). *Weighting Goods*. Oxford: Blackwell.

Dezhbakhsh, H. et al. (2003). 'Does Capital Punishment have a Deterrent Effect? New Evidence from Postmoratorium Panel Data', *American Law and Economic Review*, Vol. 5, pp. 344–376.

Donohue, J. and Wolfers, J. (2006). 'The Death Penalty: No Evidence for Deterrence', *Economists' Voice*, April.

Fagan, J. et al. (2006). 'Capital Punishment and Capital Murder: Market Share and the Deterrent Effects of Death Penalty', *Texas Law Review*, Vol. 84, pp. 1803–1867.

Foot, Ph. (1967). 'The Problem of Abortion and the Doctrine of the Double Effect', in Bonnie Steinbock and Alastair Norcross, *Killing and Letting Die. Second Edition* (New York: Fordham University Press, 1994), pp. 266–279.

GAO (General Accounting Office) (1990). *Death Penalty Sentences. Research Indicates Pattern of Racial Disparities*, available in Internet.

Griffin, J. (1996). *Value Judgement. Improving our Ethical Beliefs* (Oxford: Clarendon Press).

Harris, J. (1975). 'The Survival Lottery', *Philosophy*, Vol. 50, pp. 81–87.

Hooker, B. (2008). 'Rule Consequentialism', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/entries/consequentialism-rule/.

McCloskey, H. J. (1957). 'An Examination of Restricted Utilitarianism', *Philosophical Review*, Vol. 66, No 4, pp. 466–485.

Nozick, R. (1974). *Anarchy, State, and Utopia* (New York: Basic Books, Inc.).

Parfit, D. (1984). *Reasons and Persons* (Oxford: Clarendon Press).

Portmore, D. W. (2001). 'Can an Act-Consequentialist theory Be Agent Relative?', *American Philosophical Quarterly*, Vol. 38, No 4, pp. 363–377.

—— (2003). 'Position-Relative Consequentialism, Agent-Centered Options, and Supererogation', *Ethics*, Vol. 113, No 2, pp. 303–332.

—— (2005). 'Combining Teleological Ethics with Evaluator Relativism: A Promising Result', *Pacific Philosophical Quarterly*, Vol. 86, pp. 95–113.

Rawls, J. (1955). 'Two Concepts of Rules', in S. Freeman (ed.), *John Rawls. Collected Papers* (Cambridge (Mass.): Harvard University Press, 1999).

Scheffler, S. (1982). *The Rejection of Consequentialism. A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions* (Oxford: Clarendon Press).

Schroeder, M. (2006). 'Not so Promising After All: Evaluator-Relative Teleology and Common-Sense Morality', *Pacific Philosophical Quarterly*, Vol. 87, pp. 348–356.

—— (2007). 'Teleology, Agent-Relative Value, and "Good" ', *Ethics*, Vol. 117, No 2, pp. 265–295.

Shepherd, J. M. (2004). 'Murders of Passion, Execution Delays, and the Deterrence of Capital Punishment', *Journal of Legal Studies*, Vol. 33, pp. 283–321.

Singer, P. (1977). 'Utility and the Survival Lottery', *Philosophy*, Vol. 52, pp. 218–222.

Smart, J. J. C. (1973). 'An Outline of a System of Utilitarian Ethics', in J. J. C Smart and B. Williams, *Utilitarianism. For and Against* (Cambridge: Cambridge University Press), pp. 3–74.

Spohn, C. et al. (1996). 'The Effects of Offender and Victim Characteristics on Sexual Assault Case Processing Decisions', *Justice Quarterly*, Vol. 13, No 4, pp. 649–679.

Sprigge, T. L. S. (1965). 'A Utilitarian Reply to Dr. McCloskey', *Inquiry*, Vol. 8, pp. 264–291.

Sunstein, C. et al. (2005–2006). 'Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs', *Stanford Law Review*, Vol. 58, pp. 703–750.

Thomson, J. J. (1976). 'Killing, Letting Die, and the Trolley Problem', *The Monist* 59, pp. 204–217.

—— (1990). *The Realm of Rights* (Cambridge, Mass.: Harvard University Press).

Williams, B. (1973). 'A Critique of Utilitarianism', in J. J. C Smart and B. Williams, *Utilitarianism. For and Against* (Cambridge: Cambridge University Press), pp. 75–150.