SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI

Jacob Matthew Dragaš

ACT UTILITARIANISM VS. RULE
UTILITARIANISM

ZAVRŠNI RAD

Zagreb, 2018.
ACT UTILITARIANISM VS. RULE UTILITARIANISM

ZAVRŠNI RAD

Mentor: doc. dr. sc. Tomislav Janović
Sumentor: dr. sc. Ana Butković

Zagreb, 2018.
## Contents

Introduction.................................................................................................................................2

1. What is Ethics?..........................................................................................................................2

2. What is Utilitarianism?.............................................................................................................5

3. Act Utilitarianism....................................................................................................................7

4. Rule Utilitarianism..................................................................................................................11

5. The Intuitive and Critical Level.............................................................................................13

Conclusion......................................................................................................................................16

Bibliography..................................................................................................................................17
Introduction

This paper is a brief philosophical exploration into act utilitarianism and rule utilitarianism, both of which are two different approaches within the ethical theory of utilitarianism. I will also explore a third option which attempts to find a solution to the problems both approaches face: R. M. Hare’s differentiation between the intuitive and critical level of moral thinking. The first part of this paper defines ethics and what its main components are. The second gives a definition of utilitarianism. The third and fourth sections explore the two approaches and give examples of each. Given the brief nature of this paper, I will focus on one strength that is most difficult to give a counterargument for, and one weakness which is most problematic, for each theory. The fourth section explores a middle-ground solution. The conclusion states which of the examined theories seems most plausible while also giving a critique of the premise of utilitarianism as a whole.

1. What is Ethics?

The study of ethics is a branch of philosophy. Although philosophy does not have a universally agreed upon definition and is a topic unto itself as to whether or not it has a single definition we can pin down – for the purposes of this paper and understanding the ethical theories that will be explored, I will be defining philosophy loosely as such: the study of the nature of reality. To give an example of another great definition, Allan Buchanan of Duke University has said, “I don’t think it’s any one thing, but I think that generally it involves being critical and reflective of things most people take for granted.”

---

1 Available at http://philosophybites.com/2010/11/what-is-philosophy.html (website visited: 8th September, 2018)
One of the things one can critically reflect on is morality. How we approach morality is something we indeed often take for granted, unless we are hard-pressed to find solutions to moral dilemmas.

Suppose you are a boss at a company, and likewise suppose a very good friend of yours happens to be one of your employees. He is hard-working and has diligently performed his duties for years, in hopes of being promoted to his dream job. However, you have also been instructed by your supervisor to interview new candidates for the same position your friend is wishing to get. After the candidates have been narrowed down, the only person left is an individual who has absolutely identical qualifications as your friend. Your supervisor is adamant: he wants you to decide between the candidate and your friend and will not let you not choose. What do you do? This is an example of a moral dilemma. Either you follow the path of nepotism, choose your friend, and effectively punish the new candidate for doing nothing wrong except being good at what she does. Or on the other hand, you choose the prospective candidate and severely wound, and potentially lose, your friend knowing how much he wanted that job.

An example like this brings us to questions such as: “How do we know what is right?”, “Is there even something such asrightness?”, and “How do we define what is right?” All of these questions belong to the study of ethics. Ethics can be divided by the types of questions it is trying to answer. These include meta-ethics, normative ethics, and applied ethics. Meta-ethics concerns itself with whether or not such a thing as morality exists (is morality just a social convention or are there truly objective moral values or facts in the world independent of us?). Normative ethics attempts to find a framework, standard, or guide from which we can know what a right action is and what is a wrong one. In other words, what we ought to do, and what we ought not to do. Given this, it is a prescriptive theory, because it prescribes what action to take. And finally, applied ethics examines controversial issues such as abortion, euthanasia, and other situations.
In this work my focus will be on normative ethics, of which utilitarianism is a theory of. When examining normative theories of ethics, they can be divided generally into two groups: consequentialist and non-consequentialist theories. Non-consequential theories state that an action itself, and not its consequences, is what determines the rightness or wrongness of that action. In other words, what counts morally is what a person does, and for which reason, rather than what the results of those actions are. A classical example of a non-consequentialist normative theory is Immanuel Kant’s deontological ethics. According to Kant, not only is there a method by which we can know what the morally right action is, but we have a moral duty to choose such an action over another. Other non-consequentialist theories include theistic ethics (that our moral duties are derived from a higher power such as God, for example) and virtue ethics (that by developing good habits of character, we can live a moral life).

Let us take an example of a non-consequential normative theory, to better understand what it means and how utilitarianism will contrast it. A deontologist would typically claim that “Lying is always wrong” is a moral truth. The duty to never lie is a universal law – and as such does not depend on context, situation or one’s preferences. If you believed that such a law existed, not lying would be an objective imperative if you wanted to do what is right. To borrow Kant’s example, if a stranger came knocking on your door and asked where a family member of yours was, in order for him to kill her, would you lie? Most people would not, for lying in this case would in fact be accepted as right. Kant, however, disagreed. He believed it would be imperative to tell the truth to the stranger, even if it lead to your family member being killed, because of the universal and objective moral law that lying is always wrong. This is a classically extreme case, but most non-consequentialist theories agree that there are moral rules that should not be broken. Things like murder, rape, or theft are common ones that are cited. Of these, it is often assumed that murder is always wrong and rape is always wrong, for example.
2. What is Utilitarianism?

Utilitarianism finds itself in stark contrast with the theories discussed above because it is a consequentialist theory of ethics. Consequentialism states that an action is morally right if the consequences of that action are more favorable than unfavorable. In other words, the action itself, or the subjective motive behind it, is not what determines its rightness or wrongness, but what the effects, or consequences, of that action are. As we have seen, for non-consequentialist theories what matters is the nature of the action itself or the nature of one’s motives or reasons for that action. The consequences of actions may matter as well, but they are not crucial for their moral evaluation. Utilitarianism, however, is a theory that does not focus on rules or duties, but rather on the effects of what one does.

Utilitarianism’s origins extend back to the ancient Greek philosopher Epicurus, but today the theory is usually associated with the 18th century philosophers Jeremy Bentham and John Stuart Mill. The term “utility” was coined somewhat earlier by philosopher David Hume – using it to describe the pleasing consequences of actions as they impact people. Both Bentham and Mill used this term in describing their theories. The morally right action is the one that produces the most utility. The term “utility”, however, is a point of contention. How do we define what pleasing consequences exactly are? This was the crux of the early utilitarianism presented by Bentham and later refined by Mill.

In his book, *Ethics: A Pluralistic Approach to Moral Theory*, Lawrence Hinman defines utilitarianism’s main claim as: “The morally right action is the one that produces the greatest overall positive consequences for everyone.” Utilitarianism not only focuses on counting consequences – for utilitarians, consequences are the only things that count. What, however, are the consequences that count?

---

2 S. Nathanson, “Ethics”.
Bentham’s definition of utility centered around achieving a certain state, such as happiness or pleasure. If an action produces more pleasure, it is morally good. If it produces the opposite – pain or suffering – it is morally bad. The system he invented to help determine what our course of action in any given circumstance should be is known as the \textit{hedonic calculus}. Hinman illustrates this as follows: Using \textit{hedons} as units of pleasure or happiness, we can weigh up what brings more overall pleasure in any given situation. If for example buying your co-workers coffee produced 500 hedons of pleasure (they enjoy the coffee, you enjoy their company and the coffee) and buying it for yourself only brought 100 (let us suppose food or drink tastes better in company), you should decide on the choice that results in more overall pleasure.

The problem with Bentham’s model was that his definition of utility was overly simplistic. Take someone who makes frequent trips to Third World countries helping those in need. This person’s task isn’t necessarily pleasurable or easy – in fact it is difficult and requires a certain amount of stamina and willingness to work in conditions she is not used to. However, it would still seem that the result of this is increasing the utility of those being helped. The person volunteering also gains a happiness that isn’t solely based on a pleasure in the here and now. Higher-level concepts such as these would not pass the test of Bentham’s hedonic calculus. If we define utility as pleasure or happiness, utilitarianism cannot give an answer to living the truly good life, at least not the one we already see reflected in the desires and actions of people.

Mill refined Bentham’s theory by taking into account different kinds of happiness. He separated pleasures into two kinds. The lower, which pleasures include sex, eating, drinking, etc., and the higher: creativity, knowledge, aesthetics, etc. Mill broadened utilitarianism to include not only immediate, or solely gratifying pleasures, but results that brought about things we normally consider worthy or good but which are not tangibly pleasurable. In Mill’s case, if you were presented with helping people in a Third World country or not helping them, choosing to help them would indeed constitute a higher-level
happiness since you could predict it was a worthy investment of your time and efforts. The consequences of your actions would produce a higher level of utility overall.

So far, we have seen that utilitarianism has certain features. It is non-consequentialist (consequences, not actions are what morally counts), does not depend on rules (whatever the action is, if it results in more utility, it is morally right), and is a benevolent theory (it takes into account not only the individual, but prescribes increasing the utility for everyone). This last feature of benevolence is something that sets utilitarianism apart from competing theories, and is an important aspect in understanding its aims, and the aims of the theories we will further explore. Instead of morality solely being the responsibility of an individual, utilitarianism incorporates the condition of all human beings (or even all living beings) as a whole. Returning to Hinman’s definition above, the word “everyone” is crucial. The person making a decision should take into account a general increase and decrease of utility, not simply their own harm and benefit. A good example of this is found in Australian philosopher Peter Singer’s work. Singer is known for arguing that people living in affluent countries should not buy luxury items while there are still people living in poverty. Utilitarians agree on these features, however, what they do not agree on is answering the question: how does one decide which action is best?

3. Act Utilitarianism

Act utilitarianism offers a solution to this problem. I also think it is closer to utilitarianism’s main idea, which is that actions count. Hinman specifies its main criterion of how we should know what the right action is: “We must look at the consequences of each individual action when attempting to determine its moral worth.” In Louis Pojman’s formulation: “An act is right if and only if it results in as much good as any available alternative.” In other words, act utilitarians affirm that a case-by-case approach is the best way to assess

---

6 S. Nathanson, “Act and Rule Utilitarianism”.
any moral decision or quandary. Given that utilitarianism does not have any absolute rules, weighing the pros and cons and considering each alternative is the principle method of making sure to maximize utility for everyone.

Act utilitarianism’s greatest strength, I would say, is that of its flexibility. Earlier, we saw that a type of non-consequentialist ethics such as the deontology of Kant demands us to obey universal moral laws. Kant’s method, on how we should come to knowing what these laws are, is: “Act only on that maxim through which you can at the same time will that it should become a universal law”. In *Utilitarianism: For and Against* by J. J. C. Smart and B. Williams, Smart, a proponent of act utilitarianism, points out that his theory avoids so-called “rule worship”. Always following a rule regardless of context seems to be common sense, but as Smart puts it, “to refuse to break a generally beneficial rule in those cases in which it is not most beneficial to obey it seems irrational and to be a case of rule worship.” Rationality, therefore, is at the heart of this theory. If we are to maximize utility, why would it not be done in a manner that would be most effective, setting aside our personal biases and feelings.

I would say this is a strong case for two reasons. One, if we are able to estimate that a given choice will yield higher amounts of well-being, there is less red tape within ourselves, so to speak, to prohibit us from making that decision. In fact, there is no red tape at all. If we assume we are rational and intelligent enough to make these predictions, we can effectively lessen many results that yield less utility. Two, there is a considerable amount of freedom within this theory. If it seems there are only two options in making a decision – one that will yield less utility and one that will yield more – the likelihood of their being alternatives is fairly probable. Hence the ability to further choose which choice seems best, because it is ultimately not a matter of just two choices: obeying the rule or disobeying the rule. Richard Brandt touches on this notion in his defense of rule utilitarianism: “… it is a serious matter

---

to have a moral rule at all, for moral rules take conduct out of the realm of preference and free decisions.”

However, I would argue that as much as act utilitarianism seems at face value like a consistent approach in solving the problem of determining what actions to choose, it fails to address a critical issue of individual trust between parties. Trust is not something act utilitarianism has a lot of room for given that one ultimately can’t know the intentions of other people (assuming they are utilitarians too), other than that of them also striving for maximum utility. Here, the problem of the ends justifying the means seems to be pronounced.

To give a better illustration of this problem, let us take Christopher Nolan’s 2008 film *The Dark Knight* as an example. In the story’s conclusion, three characters play out its final scene: Batman (vigilante of Gotham city who has one rule: never kill), Harvey Dent (district Attorney of Gotham who seeks justice by fighting crime and corruption), and Jim Gordon (police commissioner of Gotham). While Batman has always been seen as a hero to many by fighting crime and stopping enemies from destroying the city, Harvey Dent has become seen as perhaps an even worthier hero because he seeks justice by working with and by the law, rather than circumventing it like Batman. He is praised by many and seen as Gotham’s answer to solving organized crime. However, at this point in the final scene, and unbeknownst to the citizens he serves, Dent has given in to his dark side after losing the person he loved, and has turned to chance and chaos to determine who of his enemies will be wiped out by whatever means necessary, even murder, and goes on a killing spree. Dent has captured Gordon’s son, and threatens to kill him assuming Gordon is the one responsible for the death of his fiancé. Before he can kill the boy, Batman tackles him, inadvertently causing Dent to fall to his death.

---

12 R. B. Brandt, “Toward a credible form of utilitarianism”, p. 158.
Gordon and Batman are now left with a choice. Either they reveal to the public Dent’s fall from grace, allowing his reputation to be destroyed, and erasing all the good he has done for Gotham, or they blame the death of Dent on Batman. Gordon and Batman both choose the latter. And I would say this choice is act utilitarian in nature. Instead of telling the truth to the public, and allowing the consequences to unfold as they will (Gotham could very well lose hope in its leaders and be in worse shape), Batman and Gordon chose a lie, tarnishing Batman’s image as Gotham’s hope, trading it in for Dent’s pure image of a “White Knight” – someone who ought not to have done wrong – for the people to place their ideals in.

In short, they chose a lie to uphold potentially better consequences, instead of the truth that would bring about potentially negative outcomes. On act utilitarianism, there is no reason to object to their action. The lie will maximize utility with the people maintaining hope, and in turn the city might rise above crime. But the problem is exactly this. Even if we assume the lie can be kept, there is no sure guarantee the results will bring about more good for more people. And if the lie is revealed, not only is hope lost in Dent – but trust is broken with Batman. We simply cannot tell what will happen in an intricately complex web of action and consequences. Yet, act utilitarianism’s premise is that of predicting what might occur and choosing that over an action one could feel is morally wrong.

My argument would therefore be in line with the critique described by Hinman that act utilitarianism opens the door for abuses since there are no rules.13 I think trust, however, as an additional component to this criticism, is act utilitarianism’s biggest weakness because it erodes the ability for us to presume others are telling the truth.

---

4. Rule Utilitarianism

The principle behind rule utilitarianism is to take the best of what utilitarianism has to offer, while proposing we should abide by rules to guide us in our moral actions. Hinman again provides a clear definition: “We ought to act in accordance with those rules that will produce the greatest overall amount of utility for society as a whole.” Rule utilitarianism is not claiming that these rules exist independently of us similar to what perhaps a non-consequentialist would claim. Instead it is stating that these rules that we ought to abide by are better guides than relying on starting from square one for each and every case. Pojman narrows the definition down: “An act is right if and only if it is required by a rule that is itself a member of a set of rules whose acceptance would lead to greater utility for society than any available alternative.” Instead of reinventing the wheel each time like act utilitarianism seems to propose, we can utilize rules as templates to re-use because we know they will bring about a greater amount of utility for everyone. I think this is a more robust approach, because it is less dependent on flexibility and more reliable in its results. The burden to deliberate on a moral action is not placed directly on the person making such a choice.

Rule utilitarianism can be compared to the rules on the road. The rules of the road are not debatable or open to personal interpretation. There are hard rules, such as the stop sign. One must obey such a rule because those are the laws that govern traffic. In a sense, we could take this as what rule utilitarianism is suggesting. The rules are to be followed because they are agreed upon to produce the highest amount of utility. A yield sign, would, for example, be more akin to what we’ve seen in act utilitarianism. One should stop only if one assesses stopping is necessary.

16 S. Nathanson, “Act and Rule Utilitarianism”.
17 S. Nathanson, “Act and Rule Utilitarianism”.
Some rule utilitarians suggest going through a number of levels when deciding whether to break a rule or not in certain cases. Pojman points out three levels: The lowest level is a set of utility-maximizing rules (e.g. “Don’t lie”), if two rules conflict in a given situation, the next level is applied.18 The next level would say that it is better to break the rule if the consequences will result in greater utility.19 And finally the third level known as the remainder rule, simply defaults to act utilitarianism because there is no other alternative.20

As already stated, I would argue the theory’s strongest feature is its robustness compared to act utilitarianism. While keeping to rules, it is a more contained approach to moral living, requiring less of the individual. It also places the rule at the top of the list to consult, rather than as a recommendation, such as “rules of thumb” (general rules that help, but do not need to be adhered to). While act utilitarianism’s strength was freedom, rule utilitarianism weighs our obligation to rules as more valuable, at the cost of lessening freedom. It also addresses my main objection of trust with act utilitarianism. In a world where everyone would agree that in general, lying is wrong (and undermines trust), there would be less fear of someone intentionally following through with such an action. Of course, I do not think this goes far enough since such a rule, in a different circumstance could potentially be broken because lying in such a situation would yield more utility.

Brandt claims that rule utilitarianism, if viewed in a more basic form, is not too different from the act utilitarianism we have so far seen. He states that, “Every act, that is to say, which maximizes utility does so because of some doubtless very complex property that it has.”21 Simply compiling a list of rules that have the property of maximizing utility, Brandt acknowledges, seems to bring us back to act utilitarianism. Hinman also points out that if a rule utilitarian does not uphold a rule in certain circumstances, it can be said that he is

21 R. B. Brandt, “Toward a credible form of utilitarianism”, p. 163.
really just an act utilitarian after all.\textsuperscript{22} Brandt’s answer to this is contained in his argument for his lengthy, but precise, definition of rule utilitarianism: “An act is right if and only if it conforms with that learnable set of rules the recognition of which as morally binding – roughly at the time of the act – by everyone in the society of the agent, except for the retention by individuals of already formed and decided moral convictions, would maximize intrinsic value.”\textsuperscript{23}

I would argue that the weakest point of rule utilitarianism is its reliance on utilitarianism itself. As we have seen earlier, the theory can ultimately collapse back into act utilitarianism. The only solution to this problem would be for rule utilitarianism to modify general rules into more specific ones. However, if it were to do this, unintended consequences could arise. For example, the general rule of “Do not lie” could be specified to “Do not lie except to prevent severe harms to people who are not unjustifiably threatening others with severe harm.”\textsuperscript{24} The problem with this rule, however, is that it would forbid lying in general, but would also “… permit lying to a murderer to prevent harm to the intended victims even if the lie would lead to harm to the murderer.”\textsuperscript{25} Not only does this allow for overly complex rules to come into play, it also undermines rule utilitarianism’s premise of offering a simpler solution to act utilitarianism.

5. The Intuitive and Critical Level

Act utilitarianism and rule utilitarianism have their strengths, but their shared weaknesses, as I have tried to suggest, is that they do not solve the main issue of utilitarianism itself: what is the process by which we can know what choice will lead to the greatest amount of utility? Act utilitarians state that the burden is on the individual to assess, and rule

\textsuperscript{23} R. B. Brandt, “Toward a credible form of utilitarianism”, p. 184.
\textsuperscript{24} S. Nathanson, “Act and Rule Utilitarianism”.
\textsuperscript{25} S. Nathanson, “Act and Rule Utilitarianism”.

13
utilitarians incorporate rules that should be adhered to as the first level of consultation. Rule utilitarianism avoids the problematic flexibility act utilitarianism possesses. And even if its proposal of adhering to rules provides a quicker, and intuitive, method of decision making (there is often less of a need to spend time deliberating on every decision), rule utilitarianism still does not give an answer to the probable possibility of itself collapsing into act utilitarianism.

Philosopher R. M. Hare proposes a third way to approach utilitarianism, which is neither of the two theories discussed in this paper so far. Hare’s solution to the problems faced by both act and rule utilitarianism lies in his presentation of a fundamental misunderstanding of two basic components of moral thinking: the intuitive and the critical level. Act utilitarianism, due to its case-by-case evaluative premise, is akin to critical thinking in its nature, but it tends to become over-specific. Rule utilitarianism, as we have seen, is more intuitive by nature, but faces the problem of collapsing into act utilitarianism. But both these theories, according to Hare, can live in conjunction if we can distinguish between these two types of thinking, since they are not competing modes of acting.

In order to illustrate this, Hare provides an eloquent example of the archangel and the prole. In this example, two extremes are imagined. On one hand is a being (“the archangel”) possessing superhuman abilities of knowledge and having no weaknesses. However, he only uses critical thinking. If he is presented with a moral decision, he will know every consequence there is to know and be able to make the best possible decision with that information. On the other hand is a person (“the prole”) with every human weakness and who does not have the ability of critical thinking and can only use his intuition. He will only know what the right actions are through the rules he has learned. Although extremes, these two sides give us an idea of why neither is better than the other.

---

26 R. M. Hare, *Moral Thinking*, p. 43.
27 R. M. Hare, *Moral Thinking*, p. 43.
28 R. M. Hare, *Moral Thinking*, p. 43.
29 R. M. Hare, *Moral Thinking*, p. 43.
Hare’s answer to the question of how we should know when to use our intuitive mode and when to use our critical one is this: “There is no philosophical answer to the question; it depends on what powers of thought and character each one of us, for the time being, thinks he possesses.” And this is a surprisingly apt answer. Hare, instead of finding solutions to both kinds of utilitarianism we have seen in this paper, is almost taking a common-sense approach, or rather not to confuse terms, a practical or tangible approach to this problem.

What makes Hare’s case strong is not only his distinction and combination of these types of moral thinking, but that balance between these two is the key to navigating the utilitarian moral life. Similar to rule utilitarianism, Hare recognizes that what we have learned contributes to the rules we follow. Learning provides an intuitive and reactionary method of action that does not require much thought. And the rules that come from intuition are sometimes more useful than that which we could arrive at solely using our critical levels of thought.

However, I would say a weakness can be found in Hare’s solution in his reframing of how we define morality. Although I think his case is stronger than that of act utilitarianism and rule utilitarianism, he seems to also be taking the problem back to the beginning. Hare explores three criteria for what one ought to do: “We might suggest as a first approximation that a use of ‘ought’ or ‘must’ is a moral use in this sense if the judgement containing it is (1) prescriptive; (2) universalizable; and (3) overriding.” I do not think his first two premises are problematic, but his emphasis on the possibility of this sense of overriding, is. He goes on to say, that “… to treat a principle as overriding, then, is to let it always override other principles when they conflict with it and, in the same way, let it override all other prescriptions, including non-universalizable ones (e.g. plain desires).” In short, when discussing his idea of combining the intuitive and the critical modes, overriding is still an option between two, similar to how we saw in rule utilitarianism. However, the reason he

---

30 R. M. Hare, *Moral Thinking*, p. 45.
31 R. M. Hare, *Moral Thinking*, p. 55.
gives for staying within utilitarianism is that the objections raised against utilitarianism often include unusual examples.\textsuperscript{32} I feel the argument I raised using Batman might be one such case Hare would not approve of. However, such an answer to the problems raised against utilitarianism is not strong enough. Unusual examples Hare mentions: “... such as the sheriff who knows – who can say how? – that the innocence of the man whom he hangs in the general interest will never be exposed” are often used to illustrate the premise of utilitarianism being faulty, and not whether or not such examples must necessarily relate to reality.

**Conclusion**

We have looked at what ethics is, what utilitarianism is, and what act utilitarianism, rule utilitarianism, and Hare’s solution is. While act utilitarianism is in a sense the utilitarian theory’s basic stance and has an advantage of flexibility of choice, it is not able to guarantee trust among individuals due to its assessing of acts on a case-by-case basis with rules playing no role. Rule utilitarianism accepted that one ought to follow agreed upon rules, but if pressed collapses back into act utilitarianism. Hare’s solution reframed the debate as the theories being complementary, rather than incompatible, and argued that these two types of prescription do not have to be opposed. I think that of these three approaches, Hare’s model provides a more complex and grounded view. However, all three seem to face the problem affecting utilitarianism in general, i.e. its dealing with consequences themselves, rather than with the motives and reasons behind actions that produce those consequences.

\textsuperscript{32} R. M. Hare, *Moral Thinking*, p. 58.
Bibliography


