

Essay

Utilitarians Should Be Theists

Peter van Inwagen

Department of Philosophy, University of Notre Dame, Notre Dame, IN 46556, USA; vaninwagen.1@nd.edu

Abstract

The moral theory *act utilitarianism* is defined in terms of the comparative amounts of pleasure acts contemplated as alternatives would yield. It is argued that act utilitarianism cannot serve as a guiding moral principle because (a) it is a plausible thesis that there is no such thing as the amount of pleasure that a contemplated act would yield, and (b) if that plausible thesis is in fact wrong and there is such a thing as the amount of pleasure a contemplated act would yield, no human being could know which act of a set of acts contemplated as alternatives would yield the greatest amount of pleasure. *Rule utilitarianism* is then defined in terms of acts performed in accordance with the set of moral rules universal adherence to which would yield the greatest amount of pleasure. It is argued that the only logically coherent way to regard utilitarianism as a guiding moral principle is to accept rule utilitarianism and to suppose that God (who would know rule utilitarianism to be the correct moral theory if it in fact is the correct moral theory and who would know the consequences of human beings' adopting any set of moral rules) has communicated to human beings a certain set of moral rules and commanded them to follow those rules.

Keywords: theism; act utilitarianism; rule utilitarianism; counterfactual conditionals; chaos

1. Problems for Act Utilitarianism

By *act utilitarianism*, I understand the following proposition:

A contemplated act is (morally) obligatory just in the case that its performance would yield a greater amount of pleasure than any possible alternative act, forbidden just in the case that its performance would yield a lesser amount of pleasure than some possible alternative act, permissible just in the case that that it is not forbidden, and “merely permissible” just in the case that it is permissible but not obligatory (or, equivalently, neither obligatory nor forbidden).¹

(If someone wishes to replace ‘pleasure’ in this definition with some other thing that falls under the artificial and abstract notion of “utility,” I have no objection. I am convinced that if any such thing were substituted for ‘pleasure’ in the arguments I am about to present, the resulting arguments would be sound.)²

By *an Act Utilitarian*, I mean a person who not only accepts the proposition I call act utilitarianism, but has also resolved to adopt that proposition as his or her sole fundamental guiding ethical principle. It seems to me to be evident that an Act Utilitarian must believe that it is possible for him or her to know (sometimes, at least) in which of the three categories *obligatory*, *forbidden*, and *merely permissible* a contemplated act would fall according to act utilitarianism, and must therefore sometimes know of two contemplated acts whether they would yield the same or different amounts of pleasure, and, if different amounts, which would yield the greater amount. (Presumably, Euthyphro, who accepted the proposition



Academic Editor: Klaus Baumann

Received: 22 July 2025

Revised: 10 June 2026

Accepted: 11 June 2026

Published: 14 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Piety is that which pleases the gods, would not have “resolved to use that proposition as his sole fundamental guiding ethical principle” if he had not believed that it was—sometimes at least—possible to know of various things that they pleased the gods.)

Act utilitarianism involves the concept of an “amount” of pleasure. I have no definition of ‘amount of pleasure’ to offer you. I will simply assume that these words make sense, or at least that sense can be made of them.

Act utilitarianism involves not only the concept of an amount of pleasure but also the concept of *the* amount of pleasure a contemplated act *would yield*. I suggest that to understand this concept, we appeal to the notion of the amount of pleasure contained in a given region of space–time, and that we identify the amount of pleasure, e.g., my raising my hand would yield, with the amount of pleasure that would be contained in the future light-cone of my raising my hand. Those who are not familiar with the notion of a light-cone may use in its place the following definition of ‘the amount of pleasure raising my hand would yield’:

The amount of pleasure raising my hand would yield is the amount of pleasure that would exist subsequently to my raising my hand—that is, the amount of pleasure that would exist in the whole (presumably infinite) future of the universe if I were to raise my hand.

The two definitions are *almost* equivalent.³ In the sequel, I will make explicit use only of the second definition of the amount of pleasure my raising my hand would yield (or, by an obvious generalization, the amount of pleasure any act would yield), despite the fact that it is not my official definition.

These three propositions follow from what we have said so far:

Act utilitarianism entails that it is obligatory for me (*now*, in *these* circumstances) to raise my hand if and only if the amount of pleasure my raising my hand would yield is greater than the amount of pleasure that any possible alternative act would yield.

Act utilitarianism entails that it is forbidden for me (*now*, in *these* circumstances) to raise my hand if and only if the amount of pleasure my raising my hand would yield is less than the amount of pleasure some alternative act would yield.

Act utilitarianism entails that it is merely permissible for me (*now*, in *these* circumstances) to raise my hand if and only if my raising my hand is neither obligatory nor forbidden.

Most persons, if they are deciding whether to adopt a moral theory as their sole fundamental guiding ethical principle, will insist that they will so honor a theory only if it is not wildly at variance with their pre-analytic moral convictions. And if the world has what we might call an Aristotelian future (namely

The earth will always exist and will always be populated by human beings and other “hedonic organisms” in much the way it is at present. Human beings and other hedonic organisms, moreover, will never be found anywhere but on the earth),

act utilitarianism will be wildly at variance with almost everyone’s pre-analytic moral convictions. For in that case the amount of pleasure that lies in the future of every possible human act will be infinite—and thus any two possible acts will yield the same amount of pleasure. And, in consequence, every possible human act will be permissible—a consequence that is certainly wildly at variance with almost everyone’s pre-analytic moral convictions.

But to say that is to do no more than to make a nice philosophical point against a position (that the world has an Aristotelian future) that is known to be wrong on scientific grounds, for the earth does not have an infinite future. It will probably be destroyed in five billion years when the sun becomes a red giant. And, in any case, life on earth cannot have a future of much more than a billion years, and the existence of pleasure certainly implies the existence of living things. Still, a billion years is a long time—something like ten thousand times the present age of our species. Even if, when I must decide whether to raise my hand or leave it unraised, there is a definite amount of pleasure that is the amount of pleasure that would be lie in the future of my raising my hand and a definite amount of pleasure that is the amount of pleasure that would lie in the future of my leaving my hand unraised, most of the times at which there will still be such a thing as pleasure are inconceivably remote, and chance and chaos will reign throughout the vast interval that separates us from the death of the last hedonic organism.⁴ Therefore, although I could reasonably infer that the amounts of pleasure that would lie in the futures of the two acts would not be exactly equal, I could reasonably infer the truth of neither of the propositions.

If I am going to raise my hand, the amount of pleasure that will exist subsequently to my raising my hand is greater than the amount of pleasure that would exist subsequently to my leaving my hand unraised.

If I am going to raise my hand, the amount of pleasure that will exist subsequently to my raising my hand is less than the amount of pleasure that would exist subsequently to my leaving my hand unraised.

I could infer neither of these things for essentially the same reason as the reason for which I could infer neither of *these* things:

If I am going to toss the quarter in my pocket exactly 100 times, the number of times it will fall “heads” is greater than the number of times the penny in my pocket would fall “heads” were I to toss it exactly 100 times.

If I am going to toss the quarter in my pocket exactly 100 times, the number of times it will fall “heads” is less than the number of times the penny in my pocket would fall “heads” were I to toss it exactly 100 times.

This was an epistemological point. It is, however, worth noting that there probably *is* no such thing as (for example) the number of times the penny in my pocket would fall heads if (contrary to fact) I were to toss it exactly 100 times—for (if I am not going to toss the penny exactly 100 times) the proposition.

If I were to toss the penny in my pocket exactly 100 times, it would fall “heads” 49 times is almost certainly false. (And the result of substituting any other number for ‘49’ in that sentence is almost certainly false.) According to David Lewis’s semantics for counterfactual conditionals (and I have no better to offer), this proposition is true if and only if

There is a possible world in which the penny is tossed 100 times and falls “heads” 49 times that is closer to actuality than is any world in which the penny is tossed 100 times and does not fall “heads” 49 times. See (Starr 2022; Lewis 1973, pp. 1–4, 13–19)

And why should that be the case? Why should there not be equally close “falls ‘heads’ 49 times” and “falls ‘heads’ 51 times” worlds? Or why not this: for every “49 world” there is a closer “51 world” and for every “51 world” there is a closer “49 world”?

But if there is no such thing as the number of times the penny in my pocket would fall heads if I were to toss it exactly 100 times, there is (surely?) no such thing as the amount of

pleasure that would exist subsequently to my leaving my hand unraised if I were to leave it unraised.⁵

It seems, therefore, that human beings are not in a position to be Act Utilitarians, since no human beings can believe—unless, perhaps, they are subject to delusions of epistemic grandeur—that it is possible for them ever to know in which of the three categories *obligatory*, *forbidden*, and *merely permissible* a contemplated act would fall according to act utilitarianism.

Our coin-tossing analogy can be used to provide a summary of the conclusion of the first part of this essay. Suppose that I must choose a coin—any coin I like from a large collection of coins—and toss it 100 times. And suppose the amount of pleasure that this act will yield is proportionate to the number of times the chosen coin falls “heads.” (So, for example, if the coin falls “heads” 50 times, that will produce twice the pleasure that its falling “heads” 25 times would produce.) But act utilitarianism provides no guidance as to which coin I shall choose because

—no number is the number of times any given coin would fall “heads” if it were tossed 100 times

—if that is wrong and every coin is such that there is a number that is the number of times it would fall “heads” if it were tossed 100 times, nevertheless: no human being could possibly *know* of any coin and any number *n* that if that coin were tossed 100 times, it would fall “heads” *n* times.

For example: No number is the number of times the quarter in my pocket would fall “heads” if it were tossed 100 times. But if that is wrong, and the quarter in my pocket would fall “heads” 46 times if it were tossed 100 times, no human being could possibly know that it would fall “heads” 46 times if it were tossed 100 times.

In the end, we must concede that Act Utilitarianism is not an acceptable moral position—for (if it does not imply that everything is permitted) anyone who accepts act utilitarianism accepts a theory that provides no moral guidance, since it is impossible for a human being to know what act utilitarianism implies about the moral status of any contemplated act.

2. Some Objections

Objection. You may have shown that there is no such thing as the *exact* amount of pleasure that would *certainly* be the result of any action a human being may be contemplating. But approximation and high probability will do for practical purposes—that is, the existence of a high probability of a certain approximate amount of pleasure resulting from an act is all that moral reasoning requires. Look, pretend for the sake of the example that amounts of pleasure can be measured in “hedons.” Suppose there is a probability of 0.98 that the result of your leaving your hand unraised will be within 1000 hedons of 611525724270000 hedons, and a probability of 0.98 that the result of your raising your hand will be within 1000 hedons of 2533144660000 hedons. (Compare: if one is about to toss a coin exactly 100 times, the probability that the number of times it will fall ‘heads’ lies in the range 45–55 is about 0.7286.) The 15-digit number is almost 250 times greater than the 13-digit number. Obviously, any thesis that could reasonably be called ‘act utilitarianism’ ought to imply that (in the case imagined) raising your hand is forbidden.”

I do not think this argument succeeds—but I do think that it is plausible enough that it merits careful consideration.

Suppose that there will be pleasure for the next one billion years and none thereafter. Suppose that every act contemplated by human beings living today will (to a very high probability) yield an amount of pleasure between one million plus 1000 and one million minus 1000 times the amount of pleasure experienced by all human beings born in or before the 21st century. Does that imply that, for every future act x and every future act y , the amount of pleasure that will result from x is approximately equal to the amount of pleasure that will result from y ? If it does, act utilitarianism (modified) implies that “everything is permitted.” And one might suppose that those amounts of pleasure are indeed approximately equal, since (for example) $10^6 - 611$ is about 99.887 percent of $10^6 + 520$. But it is plausible to suppose that they are very far from being approximately equal. After all, the difference between one million plus 520 times the amount of pleasure experienced by all human beings born in or before the 21st century and one million minus 611 times the amount of pleasure experienced by all human beings born in or before the 21st century is 1131 times the amount of pleasure experienced by all human beings born in or before the 21st century—a significant difference, surely?

In any case, those two amounts of pleasure are approximately equal or they are not. If they are, then it would seem that the amount of pleasure that *any* act would yield is approximately equal to the amount of pleasure any other act would yield. After all, a billion years is, as I said, a long time, and things tend to even out in the long run. Act Utilitarians who believe both that “approximation will do for practical purposes” and that the amount of pleasure any act would yield is approximately equal to the amount of pleasure any other act would yield must say “All things are permitted.” Act Utilitarians who believe that the amounts of pleasure some acts would yield are significantly different from the amounts some possible alternatives to those acts would yield will find themselves bereft of moral guidance, for it will never be possible for them to know of any contemplated act whether it has that feature, much less know whether the amount of pleasure it would yield is greater than, less than, or the same as, the amount of pleasure that some given alternative act would yield.

Objection. Your argument assumes that ‘the amount of pleasure a possible act yields’ depends on episodes of pleasure that will occur in the far future (or would occur in the far future if they did occur), even hundreds of millions of years in the future. But utilitarianism is not concerned with temporally remote and unknowable events; it is concerned the probable causal consequences of one’s acts—and events that are the probable causal consequences of one’s acts are usually events in the near future.

This is Bentham’s notion of “propinquity or remoteness,” one of the pairs of oppositions that figure in his felicific calculus (Bentham 1780, chap. IV). *Why* it is one of them is a puzzle; however, for any useful function it might have is better performed by the pair called “certainty or uncertainty.”

Let us suppose that there really are such things as the amount of pleasure that would exist subsequently to my raising my hand if I raised my hand and the amount of pleasure that would exist subsequently to my leaving my hand unraised if I left my hand unraised. And let us suppose that—in the distant land in which I dwell—whenever agents deliberate between two alternative courses of action, God tells them things like this: “The amount of pleasure that would exist subsequently to your leaving your hand unraised is 23.682 times the amount of pleasure that would exist subsequently to your raising your hand.” It seems to me to be evident that any moral theory that could reasonably be called a species of utilitarianism should imply that an agent vouchsafed that particular revelation is morally obliged to leave his or her hand unraised. That is, if (in Bentham’s terms) certainty has been secured, remoteness is of no account.

Objection. Even if Act Utilitarians face all the difficulties you say they face, you have said nothing that implies that those whose sole fundamental guiding ethical principle is *rule* utilitarianism face any difficulties.⁶

Let us say that rule utilitarianism is this twofold thesis:

A contemplated act is obligatory just in the case that failure to perform it would violate the moral rules adoption of which would have the best consequences, forbidden just in the case that performing it would violate the moral rules adoption of which would have the best consequences, and merely permissible just in the case that it is neither obligatory nor forbidden.

Certain moral rules are the moral rules adoption of which would have the best consequences just in the case that for any time t , and any possible worlds w and w' , if w and w' are alike at and before t , and if in w human beings begin at t universally to follow those rules and always follow them thereafter, and if in w' human beings begin at t universally to follow some other rules and always follow them thereafter, the amount of pleasure contained in w is greater than the amount of pleasure contained in w' .

We may define a Rule Utilitarian as someone who both accepts this twofold proposition and has also resolved to adopt it as his or her sole fundamental guiding ethical principle.

It seems to me that Rule Utilitarians face problems that are of very much the same sort as the problems that Act Utilitarians face.

Let us say that a set of moral rules A is *hedonically superior* to a set B if and only if the members of any A -Society (any society all of whose members strictly adhere to A) experience more pleasure (individually, not collectively) than the members of any B -Society. And let us suppose that there are certain moral rules—let them be called ‘the Seven Rules’—that are hedonically superior to any other rules.

Does the counterfactual conditional

Strict adherence to the Seven Rules by all human beings would yield a greater amount of pleasure than would strict adherence to any other set of rules by all human beings.

follow from that proposition? I see no reason to suppose it does. Not, certainly, if there will be sentient organisms for a billion years and sapient organisms like ourselves will exist for only a minute fraction of that vast expanse of time. One might as well maintain that if beavers built a dam at a certain spot, the average depth of the waters of the world would be increased. And there are not only vast expanses of time to consider, but the fact that the temporal evolution of the world is almost certainly chaotic: small changes in the present will snowball into greater changes in the future—into increasingly greater changes as time passes. (Compare the oft-repeated statement that a butterfly’s flapping its wings may lead to a tornado a few weeks later and a continent away.)

David Lewis (I remember) once made this point vividly in a lecture. He held up a hand and said, “Observe! I wave my hand.” [He waved his hand] “*Empires* that would have existed fifty thousand years from now if I had not waved my hand, now will not exist.” And this statement was highly plausible—if only because it is highly plausible to suppose that if *anyone* had behaved differently on *any* occasion from the way he actually behaved on that occasion, the whole human condition 50,000 years from now would be radically different from the way it would have been. (Supposing, of course, that our species has that span of time left to it.)

Whatever may be the case as regards the air in the vicinity of a butterfly’s wings, the world is full of demonstrably chaotic systems—systems in which minuscule changes

lead to very large changes very quickly. A standard example is the movement of balls on a billiards table.⁷ It is because the world contains many billiards tables—and many other equally chaotic systems. (Wikipedia gives these examples: fluid flow, heartbeat irregularities, weather, and climate.) It is because the physical world is saturated with chaotic systems that David Lewis's thesis about future empires (or their absence) seems to me to be very likely to be true.

Suppose, then, that our species has 100,000 years left to it and that there will be hedonic organisms for a billion years (ten thousand times 100,000 years). I see no reason to suppose that perfect adherence to the Seven Rules by all human beings for the next thousand centuries would yield a greater amount of pleasure than would perfect adherence to any other set of rules. In fact, suppose there is a set of moral rules that is hedonically *inferior* to all other sets of rules. Call the members of this set the Thirteen Rules. Because the world is a chaos, and because there will be sentient organisms long after there are sapient organisms, the following three counterfactual conditionals are almost certainly false:

Perfect adherence to the Seven Rules by all human beings would yield a greater amount of pleasure than would perfect adherence to the Thirteen Rules by all human beings

Perfect adherence to the Seven Rules by all human beings would yield a lesser amount of pleasure than would perfect adherence to the Thirteen Rules by all human beings

Perfect adherence to the Seven Rules by all human beings would yield the same amount of pleasure as would perfect adherence to the Thirteen Rules by all human beings.

3. Theism to the Rescue!

To the rescue of the Rule Utilitarian, that is. For theism cannot rescue the Act Utilitarian. It might have been able to effect such a rescue if (as in a thought experiment I made use of above) God were to whisper in each moral agent's ear the amounts of pleasure each contemplated course of action would yield. As we have seen, however, it is likely (on *a priori* grounds) to be false that *any* amount of pleasure is the amount of pleasure my raising my hand would yield (or the amount that any possible alternative act would yield). In any case, God has not vouchsafed us such revelations.

But a theist can be a Rule Utilitarian. We can, for example, imagine a theist who reasons as follows.

I believe that rule utilitarianism is the correct moral theory. And I believe that God exists and that he has given a certain set of moral rules, the Decalogue⁸, to humanity—those ten rules and no others. Now since (as I believe) rule utilitarianism *is* the correct moral theory, it must be that God knows that rule utilitarianism is the correct moral theory. And since he has given that particular set of rules, the Decalogue, to humanity, and since he knows that rule utilitarianism is the correct moral theory, it must be that he (who is omniscient) knows that the Decalogue is the set of moral rules adoption of which would have the best consequences. Therefore, I will regard an act as obligatory if and only if it is commanded by the Decalogue and forbidden if and only if the Decalogue commands me not to perform it.

In sum:

To be a utilitarian is to accept either act utilitarianism or rule utilitarianism. It is a consequence of act utilitarianism that one can never know which of the categories *obligatory*, *forbidden*, and *merely permissible* a contemplated act falls, and rule utilitarianism conjoined

with atheism shares that consequence. But the conjunction of rule utilitarianism and the proposition that God exists and has given humanity a set of moral rules implies that it is possible to know which of these categories contemplated acts fall in. Therefore, anyone who has resolved to adopt utilitarianism as his or her sole fundamental guiding ethical principle should be a theist.

4. Envoi

If anyone supposes that, theist that I am, I have presented an argument for the conclusion ‘Utilitarians should be theists’ in order to provide some sort of support for theism, they are mistaken. For I find utilitarianism wholly implausible (for reasons having nothing to do with considerations of the sorts presented in this essay). To borrow a figure of Quine’s, I am in the position of a Jewish chef preparing ham for a Gentile wedding.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

Notes

- ¹ For an account of act utilitarianism, see (Sinnott-Armstrong 2023). It would be advisable to read the whole article.
- ² I will not, beyond this note, mention pain. Let us think of amounts of pain as, so to speak, negative amounts of pleasure: the pains contained in a region of spacetime reduce the amount of pleasure contained in that region. Suppose, for example, that *R1* and *R2* are regions of spacetime that are perfect duplicates in respect of episodes of pleasure; and suppose that *R2* contains episodes of pain and *R1* contains none. Then the amount of pleasure in *R1* is greater than the amount of pleasure in *R2*—and this despite their being “pleasure duplicates.” An immediate consequence of this convention is that the amount of pleasure in a region can be negative.
- ³ The future pleasures of distant alien beings might well fall outside the light-cone of my raising my hand. But if I am trying to decide whether to do *A* or do *B*, the future pleasures of distant alien beings can reasonably be presumed to be the same whether I do *A* or do *B*.
- ⁴ As the Preacher has said, I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all. (Ecclesiastes 9:11)
- ⁵ I am of the view that even if *am* going to raise my hand, there is no such thing as the amount of pleasure that will exist subsequently to my raising my hand, but I will not pursue this point.
- ⁶ For an account of rule utilitarianism see (Sinnott-Armstrong 2023).
- ⁷ “A simple example serves to illustrate just how sensitive some physical systems are to external influences. Imagine a game of billiards, somewhat idealized so that the balls move across the table and collide with a negligible loss of energy. With a single shot the billiard player sends the collection of balls into a protracted series of collisions. The player naturally wants to know the effects of the shot. For how long could a player with perfect control over his or her stroke predict the cue ball’s trajectory? If the player ignored an effect even as minuscule as the gravitational attraction of an electron at the edge of the galaxy, the prediction would become wrong after one minute!” (Crutchfield et al. 1986. The quoted passage occurs on pp. 48–49).
- ⁸ Or whatever. Feel free to substitute any other set of rules for the Decalogue in this passage.

References

- Bentham, Jeremy. 1780. An Introduction to the Principles of Morals and Legislation. Available online: https://en.wikisource.org/wiki/An_Introduction_to_the_Principles_of_Morals_and_Legislation (accessed on 22 June 2025).
- Crutchfield, James P., J. Doyne Farmer, Norman H. Packard, and Robert S. Shaw. 1986. Chaos. *Scientific American* 254: 46–57. [CrossRef]
- Lewis, David. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.

- Sinnott-Armstrong, Walter. 2023. Consequentialism. In *The Stanford Encyclopedia of Philosophy*, Winter 2023 ed. Edited by Edward N. Zalta and Uri Nodelman. Stanford: Metaphysics Research Lab. Available online: <https://plato.stanford.edu/archives/win2023/entries/consequentialism/> (accessed on 22 June 2025).
- Starr, Willow. 2022. Counterfactuals. In *The Stanford Encyclopedia of Philosophy*, Winter 2022 ed. Edited by Edward N. Zalta and Uri Nodelman. Stanford: Metaphysics Research Lab. Available online: <https://plato.stanford.edu/archives/win2022/entries/counterfactuals/> (accessed on 22 June 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.