

Journal of Philosophy, Inc.

Rule-Utilitarianism and the Slippery Slope

Author(s): Gregory W. Trianosky

Source: *The Journal of Philosophy*, Vol. 75, No. 8 (Aug., 1978), pp. 414-424

Published by: [Journal of Philosophy, Inc.](#)

Stable URL: <http://www.jstor.org/stable/2025565>

Accessed: 07/09/2011 01:59

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Journal of Philosophy, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*.

<http://www.jstor.org>

RULE-UTILITARIANISM AND THE SLIPPERY SLOPE *

IT is sometimes said that permitting some form of euthanasia would gradually erode moral motivations and behavioral inhibitions that should support a moral code. It is said, for instance, that permitting voluntary euthanasia would lead to erosion of inhibitions about acts of involuntary euthanasia, which in turn would erode motivations and inhibitions on killing in general to the point where we would wink at euthanasia for those who are a nuisance to society: idiots, recidivist criminals, defective newborns, and the insane, for example.¹

Arguments of this sort appear constantly in the philosophical, theological, legal, and popular literature on contemporary moral problems. The debates on euthanasia, abortion, fetal research, and suicide are full of them. They appear under a variety of titles, but I will call them "slippery-slope" arguments; for the situation they envisage is an inevitable and progressive deterioration of motivations and inhibitions—a slide pell-mell down the slippery slope—once we take the first step by releasing our hold on a strict moral code and accepting instead some weaker set of moral rules and prohibitions.

In this paper I want to explore the relationship between certain of these arguments and the kind of normative ethical theory exemplified by ideal-rule-utilitarianism. I will begin by explaining the logic of the sort of slope argument with which I am concerned, and of this sort of normative theory. Then I will argue that such theories cannot properly take into account the force of bona fide arguments of the slippery-slope genre. Finally, I will tentatively suggest a related but importantly different sort of ethical theory which can give due weight to such arguments. I will take the slippery-slope arguments on voluntary euthanasia as typical examples.

I. THE SLIPPERY-SLOPE ARGUMENT

Undergirding any accepted moral code is a vast web of motivations, behavioral inhibitions, attitudes, and beliefs. A large portion of these should support prohibitions and obligatory limitations on killing. Thus, in virtue of accepting such a code, we should be mo-

* I want to thank Holly S. Goldman for her invaluable advice and assistance. A version of this paper was read at the meetings of the Western Division of the APA in April 1978.

¹ This particular formulation of the argument comes from an unpublished paper of Tom L. Beauchamp's entitled "A Defense of the Distinction between Active and Passive Euthanasia."

tivated not to take the lives of others; we should disapprove strongly or even blame those who do; we ourselves should experience remorse, regret, or perhaps guilt if we have killed or intended to kill someone; and so on. I will later say (sec. II), with Brandt and probably Hare, that to have these and similar sorts of feelings and motives is *part of what it means* to accept a moral code that includes obligatory limitations and prohibitions on killing.

The slippery-slope argument with which I am concerned claims that, as a matter of psychological fact, if our moral code included some qualifications or relaxations of our obligations with respect to killing, our inhibitions regarding killing in general would be weakened to such an extent that significant disregard of the moral law would result. And, though many slope claims are wildly implausible, it seems to me that there is an important core of truth in such arguments. The various portions of the supporting psychological web of motives and inhibitions are no doubt mutually interdependent; so a code characterized by the qualification or absence of certain central provisions (e.g., those governing obligations on killing) might well be a code that weakened the supporting motives and attitudes for related prohibitions. I think this is because the ordinary psychological techniques that would be most successful in reinforcing a permissive attitude toward (say) voluntary euthanasia are too clumsy to be subject-specific: they cannot effectively be directed toward only *one* qualification of the prohibition against killing; and the reinforcement of permissive attitudes in one area will thus "spill over" from its target and alter other related motivations and attitudes as well.

The slippery-slope argument discussed here then makes the following claim: no action can be right, other things being equal, if its being permitted by the moral code would so seriously undermine the reinforcement of other prohibitions and constraints as to outweigh any benefits that would be gained from permitting it. It is upon this claim, often implicit, that the weight of the slippery-slope argument under consideration rests.²

² Some philosophers may be concerned with the problem of conceptual vagueness and the slippery slope. Compare, e.g., Norman C. Gillespie's recent paper on abortion in *Ethics*, "Abortion and Human Rights," LXXXVII, 3 (April 1977): 237-243. I will not discuss this sort of slope argument at all. Moreover, as the text suggests, I am not interested in the "slope" problems that might arise in the *transition* from accepting one moral code to accepting a different one. The slope arguments with which I am concerned focus rather on *internal* strains generated by the conflict of provisions within the code itself, or after it begins to operate.

II. IDEAL-RULE-UTILITARIANISM

I have already said a little about the notion of *acceptance* of a moral code; now I will sketch more fully the kind of theory based on this notion with which I will be concerned.

An ideal-rule-utilitarian might formulate his ultimate answer to the normative question, "What is it right for me to do?" in roughly the following way:

- (1) An act is right if and only if it is permitted or recommended by the moral code whose acceptance in the agent's society would maximize utility.

There are a number of ways in which the notion of acceptance might be spelled out; but I will focus on what I think is the most plausible sort of analysis, following Richard B. Brandt and R. M. Hare:

- (2) An agent accepts a moral code only if
- (A) He is motivated to bring about or to avoid bringing about the states of affairs recommended or proscribed by the code; e.g., he is motivated not to cause injury to others.
 - (B) If he is aware of having done one of the things he is so motivated not to do, then (unless he believes he is excused) he will feel uncomfortable (or, in some sorts of cases, guilty) to a degree roughly commensurate with the strength of the motivation to do or avoid.
 - (C) Normally, he is disturbed by the presence in anyone else of the conduct he is so motivated to avoid; and the disturbance may reach the point where he is motivated to express it in the form of criticism (or in some sorts of cases, blame).³

If we are thinking of duty ethics codes, then paramount among the states of affairs the code recommends, and hence among those which the agent who accepts the code must be motivated to bring about, will be states of affairs like "the truth's having been told," or "a promise's having been kept." The agent who accepts a duty ethics code will necessarily, by (2), be motivated to act and to approve of acting in these ways recommended by the code.

³ Brandt, "Utilitarianism," John Locke Lectures, delivered at Oxford, 1974 (unpublished). Cf. also his "A Utilitarian Theory of Excuses," *Philosophical Review*, LXXVIII, 3 (July 1969): 327-361.

Hare's view is nowhere stated in this detailed way; but it seems to agree in broad outline with the view I describe in (1) and (2). Cf. his remarks in "Principles," the Presidential Address in the *Proceedings of the Aristotelian Society*, LXXIII (1972/73): 1-19, pp. 12, 16/7, regarding inculcation and acceptance.

The rule-utilitarian sort of normative ethical theory which I have formulated in (1), supported by the partial analysis of acceptance provided in (2), is only one of a family of views which we can call *code-acceptance theories of rightness*.⁴ So, a deontologist might hold something quite like (1) in spirit, namely:

- (3) An act is right if and only if it is permitted or required by the moral code whose acceptance in the agent's society would result in some deontological principles *D* and *E* having been acted on to the greatest degree possible.

The crude version of ideal-rule-utilitarianism in (1) thus suggests a whole group or family of code-acceptance theories: theories, that is, which take an action to be right just in case it conforms to a code whose acceptance would bring it about that some morally desirable state of affairs⁵ obtained to the greatest extent possible, whether it be "utility's having been maximized," or "there being fair and just distribution of certain goods," or "courageous actions having been performed," or some combination of these, or something else.⁶ I have no particular axe to grind in favor of utilitarianism, and the issue of the relation between ideal-rule-utilitarianism and the impact of slippery-slope arguments clearly has much wider application and interest than its formulation in utilitarian terms might suggest. I will take the utilitarian code-acceptance theory to be illustrative of the issues under discussion.

⁴ Note that rule-utilitarianism as described in (1) is popularly taken to be a normative ethical view and not a metaethical view; and I will so understand it. Cf. Brandt's remarks in "Some Merits of One Form of Rule-Utilitarianism," reprinted in Kenneth Pabel and Marvin Schiller, eds., *Readings in Contemporary Ethical Theory* (Englewood Cliffs, N.J.: Prentice-Hall, 1970), pp. 282/3. Of course the analysis of the notion of acceptance in (2) may embody or constitute a metaethical view.

⁵ 'Morally desirable states of affairs' is a catch-all phrase. Some states of affairs are morally desirable because they are good in a nonmoral sense—e.g., "utility's having been maximized." Others are not necessarily held to be good in this way—e.g., "fair and just distribution of goodness obtaining."

⁶ I take Robert Nozick to have shown, in his discussion of "rights-utilitarianism," that not all theories of the right can be translated into maximalist views; see *Anarchy, State and Utopia* (New York: Basic Books, 1974), pp. 28–32. But it does not follow from this that my claim here is false, unless this shows that not all theories of the right can provide *some* preference-ordering of possible worlds, from the world in which (e.g.) everyone's rights are always respected to that in which no one's rights are respected. And it seems to me that Nozick has not shown this, but rather only that not all theories of the right will rely on some sort of maximizing criterion to establish this ordering. This issue needs to be discussed at greater length, however.

III. THE ISSUE

Let us suppose, then, that a code-acceptance utilitarian is confronted with a typical slippery-slope argument. A more careful statement of the argument might go: "If we were to accept a moral code which, while prohibiting almost all other forms of killing, permitted voluntary euthanasia, then the motivations and attitudes that supported the general prohibition on killing—or at least particular portions of it—would slowly but surely be weakened or destroyed. Hence utility would not be maximized by such a code's acceptance."

We can simplify the argument as follows. Assume a simple moral code *C* which deals only with killing and which has only three parts: *K*, *L*, and *M*. *K* is the set of rules permitting voluntary euthanasia; *L* is the set of rules on killing for which motivations and attitudes are endangered; and *M* is the set of remaining (undamaged) killing-rules. Then the argument runs:

- (4) If a code *C* is accepted with provision *K* in it, then the motivations and attitudes that support *L* will be weakened or destroyed. Hence utility would not be maximized by our accepting *C*.

What impact does this sort of argument have on the rule-utilitarian view as stated in (1)? What effect can it have on our decision, so based, whether or not to adopt the code? *Absolutely none.*

The rule-utilitarian who defends a crude code-acceptance view like (1) is concerned only with the question of whether utility would be maximized *if the entire code were accepted*. He thus *assumes* that the code would be accepted exactly as written and then asks whether or not, if that were so, utility would be maximized. But, as the partial definition of acceptance on which we rely in (2) makes clear, if code *C* were accepted in toto by an agent, then a fortiori the agent would have the requisite motivations and attitudes supporting provision *L*, as well as those supporting *K* and *M*. Thus, if the antecedent of the first sentence of (4) is true, its consequent must be false in virtue of the definition of acceptance. It cannot be the case both that an agent accepts *C*, in our sense, and yet that he lacks the motivations and attitudes that support each and every one of the provisions of *C*.

In short, the slippery-slope argument in (4) has no impact, if the code-acceptance line embodied in (1) is taken. For what we then do to pick a moral code is to construct an entire code, assume that it is accepted, and *then* ask about its utility. And if such a code

were accepted, then, by definition, agents would be motivated to obey all its provisions.

Now I have been discussing an analysis on which the *defining characteristic* of acceptance is essentially the agent's possession of a certain *level of motivation*. There is a more standard analysis of acceptance, however (familiar from the equivalence-thesis debates), on which a certain fixed *frequency of performance* is taken to be the defining characteristic.

But this standard, *actional* analysis even more clearly exhibits the same blindness to slippery-slope considerations as the *motivational* analysis under discussion. The payoff of slope arguments, after all, comes in the claim that, because motivations are adversely affected by the psychological tensions generated by certain combinations of rules, the result will be widespread disregard of or failure to *act* on various provisions of the moral law. But, if we assess a code solely by reference to the consequences of its acceptance, *and the defining characteristic of acceptance is a fixed, relatively high frequency of performance*, slope arguments will clearly have no impact. For if a code is accepted, on this more standard view, agents will by definition frequently or usually act as it requires; and our calculation of the code's utility will thus *assume* the frequent or usual performance of all recommended actions.

IV. WAYS OUT?

I believe that slippery-slope claims should be important considerations which can weigh against our choice of a moral code and, hence, which must (on the rule-utilitarian view) be relevant to the question of the rightness of any particular sort of action. There is an important psychological insight in the claim that certain combinations or sets of rules might generate internal stresses or conflicts in moral agents and that such psychological stresses might seriously affect the probabilities of their acting as the code required or recommended. Moreover, the inability of code-acceptance theories to deal with slope arguments is just a special case of their inability to consider a whole range of important questions about possible psychological conflicts regarding moral rules. So, for instance, there is the possibility of motivational tensions or stresses between the provisions of the code and the concerns of self-interest, or between what morality requires and what our concern for loved ones moves us to do. All these sorts of psychological conflict, and their impact on the probability of the performance of right action, *must be disregarded* by the code-acceptance theorist in his or her assessment of competing moral codes.

And this seems to me a devastating difficulty for such theories. I cannot see that we would necessarily have good reasons to act on—or even to press for the acceptance of—the recommendations of a code which, though it would maximize utility if accepted in toto, cannot be so accepted because of important psychological facts about the relations among various motives, attitudes, and beliefs of moral agents.

If it is indeed correct, therefore, that slippery-slope arguments like (4) carry no weight against views like (1)—given either the motivational or the actional analysis of acceptance—then it seems to me that the entire value of the rule-utilitarian sort of project of providing criteria for the rightness of action and for the justification of normative principles by analyzing the consequences of a code's acceptance is called into serious question.

Can the code-acceptance view in (1) be amended or clarified so as to allow slippery-slope claims as a relevant consideration for our choice of a moral code and, hence, as relevant with respect to the rightness of actions? I will now consider one possible clarification and two possible emendations of the crude code-acceptance theory exemplified by (1). I refer for simplicity only to the motivational analysis with which we began, although my remarks should apply *mutatis mutandis* to the more standard actional analyses as well.

A Clarification. It might be pointed out that to accept a code is indeed (in part) to be motivated in certain ways; but acceptance requires only *some degree or other* of motivation. Motivations are not “all or nothing.” Two agents who both accept code *C* may have substantially different *degrees* of motivation, and nothing in the code-acceptance view denies this. But then slippery-slope arguments to the effect that the degree of motivation will be *weaker* if *C* is accepted than if another code without provision *K* were accepted, do have some weight and are relevant considerations in our choice of a moral code.

I find this clarification of the code-acceptance theory unconvincing in the present case.

(i) It has unnatural and paradoxical results. On this view, any slippery-slope argument that claims merely that, if code *C* were adopted, motivations and inhibitions supporting provision *L* will be weakened, does carry weight. But any slippery-slope argument that makes the stronger claim that adoption of *C* would altogether destroy the motivations supporting *L* would still carry no weight at all against (1). The clarification permits us to give weight to the

weaker sorts of slope arguments, but still requires us to ignore the stronger ones.

(ii) Moreover, it seems to me that acceptance of a code on the motivational analysis must surely entail some minimal degree of motivation such that the moral motivations of the accepting agent exert significant and frequently decisive influence on the outcome of his or her deliberations about what to do.⁷ But the real force of many slope arguments, as I've noted, comes precisely in the claim that motivations will be so substantially weakened that in a great number of cases they will no longer exert a significant or decisive influence on action at all. And this, I would think, would not be compatible with the acceptance of a code, even on the motivational analysis. If this is so, then *any* slope argument on the weakening of motivations which has real force is still removed from consideration by the code-acceptance theory as clarified.

Two Emendations:

(i) *The Learnability Requirement.* Brandt does not subscribe to the crude version of a code-acceptance theory stated in (1), as even a cursory examination of his earlier writings will indicate. Nor, seemingly, does Hare.⁸

We might phrase one relevant additional condition that they impose in the following way:

(5) The code must be such that its acceptance is *educationally possible*.

That is, not only must the code be such that its acceptance would maximize utility; it must be such that most or all moral agents of average intelligence are able to learn it.

This provision rules out codes that are too complex, as both Brandt and Hare observe; codes that contain too many exceptions and qualifications, for example. It thus places at least a bottom limit on the irrelevance of slippery-slope arguments of certain sorts. For one thing that extreme forms of such arguments might be taken to show is that certain very complex codes break down because ordinary people just can't keep all the distinctions, caveats, and exceptions straight in their heads. But the sort of slippery-

⁷ The two analyses of acceptance are thus obviously related. However, that the accepting agent generally or frequently acts as the code recommends is *entailed* by his acceptance on the actional analysis; while such frequent performance is only *causally* related to acceptance on the motivational analysis.

⁸ "Toward a Credible Form of Utilitarianism" in Hector-Neri Castañeda and George Nakhnikian, eds., *Morality and the Language of Conduct* (Detroit: Wayne State UP, 1965), pp. 123 f. Cf. also Hare's "Principles," *op. cit.*

slope argument with which we are concerned does not claim that agents will be unable to draw the relevant sorts of moral distinctions and to generate moral judgments in accordance with them. It points only to the fact that, though agents may learn the rules, they very likely just won't be strongly motivated to obey some of them.

(ii) *The Psychological-possibility Requirement.* John Rawls suggests another requirement,⁹ one version of which I think would permit slippery-slope claims as relevant considerations in our choice of a moral code. I call this version the "psychological-possibility" requirement:

(6) The code must be such that its acceptance is *psychologically possible*.

That is, not only must (1) and (5) be satisfied, but the code must be such that its acceptance would not generate motivational and attitudinal conflicts such that the acceptance of one portion of it (*K*) reduces or excludes the psychological possibility of acceptance of other portions of it (*L*).¹⁰

This additional requirement pretty clearly permits us to weigh slippery-slope claims like (4)! In fact, if such a claim is successful, then the proposed code will fail altogether to pass the psychological-possibility requirement. But this raises a new problem.

V. AN OBJECTION TO THE "WAY OUT?"

As stated, the code-acceptance theory now involves three distinct requirements, satisfaction of each of which is necessary for the ideal moral code. Thus it is a sufficient condition for rejecting a proposed code that it fail to satisfy any one of them.

But this consequence is unacceptable. Surely we want slippery-slope claims to be a *relevant* consideration in our choice of a moral code; but the present requirement structure ensures that a successful slope argument is not only relevant but necessarily *decisive*. Yet it seems to me that we might sometimes have conclusive reasons in favor of picking a particular moral code in spite of the psychological or indeed learning difficulties inherent in obtaining its widespread acceptance, and perhaps in spite of the disutility incurred by the weakening or destruction of the motivations and inhibitions involved in accepting certain of its provisions. These are factors that must be weighed in the choice of a moral code,

⁹ *A Theory of Justice* (Cambridge, Mass.: Harvard, 1971), sec. 69. Cf. G. E. Moore, *Principia Ethica* (New York: Cambridge, 1959), pp. 160/1.

¹⁰ Hare might well accept this requirement, although this is unclear. Brandt does not, at least not explicitly.

but it is at least conceivable that they could be outweighed by other considerations in favor of the code. So, for example, it might maximize utility to pick a code some of whose action-guiding provisions or rules were psychologically impossible for the average person to accept: those recommending (but not requiring) desirable but supererogatory, superhuman acts of sacrifice or perseverance, for instance. Even if such recommendations were widely disregarded, the frequency of the recommended behavior might nonetheless be higher than if there were no provision at all in the code. Or utility might be maximized by a code acceptance of some of whose provisions weakened other motivations and inhibitions, if the degree to which they did so was small, or if the weakened motivations were of relatively minor importance (motivations to do favors or minor acts of courtesy, for instance).

Similarly, we might also want to pick a code that was not learnable in its entirety by everyone. It is useful in this connection to think of the rules of golf or bridge. Everyone can play these games, and must follow the basic rules; but it is no argument against the more complex rules (used in tournament play) that most ordinary players couldn't possibly learn them. This is especially plausible if morality, like golf, involves a sort of "excellent activity"¹¹ and, perhaps like golf, not one in which all can perfect themselves.¹²

VI. A WAY OUT

Perhaps, then, we're on altogether the wrong track to add nonconsequentialist requirements to (1) as additional criteria for picking a moral code. Should we instead reject the whole code-acceptance line as exemplified in (1)? I think so.

The fundamental problem with the notion of acceptance is that it is a *success notion*: Built into it is the assumption that moral agents *succeed* in displaying certain fairly high levels of motivation (or performance) with respect to each provision of a given code. We are thus unable to recognize and evaluate internal psychological conflicts within the code, or between the code and other of the agent's (nonmoral) motivations and attitudes which may in fact preclude success. Because acceptance is a success notion, we are forced instead to *assume* that each agent will successfully maintain the requisite levels of moral motivation.

If we are to preserve the possibility of an ideal-rule-utilitarian

¹¹ For an interesting contemporary discussion of morality as involving a form of excellent activity, cf. W. K. Frankena's *Three Questions about Morality: The Carus Lectures for 1974*, soon to be published by Open Court.

¹² This shows that the conclusion of (4) is formulated too strongly.

sort of normative theory, therefore, we must replace the notion of acceptance in (1) with something else that isn't a success notion. We might for instance want to evaluate the consequences of a code's being *supported* or *promulgated* by the moral agents in our society, where promulgation might be defined as the systematic utilization of certain psychological tools for the purpose of inculcating and maintaining acceptance, here without the assumption that such efforts would necessarily succeed. But this line may have difficulties of its own.¹³

Whatever notion we choose as a replacement, however, it is clear that the code-acceptance line embodied in standard formulations of ideal-rule-utilitarianism will not do, if we are to take seriously those sorts of argument from psychological conflict of which certain slippery-slope claims are the most well-known instances.

GREGORY W. TRIANOSKY

University of Michigan

COMMENTS AND CRITICISM

TOWARD A CAUSAL THEORY OF EVIDENCE *

W. V. QUINE's program in epistemology and semantics arises from what Geoffrey Joseph has termed a "Humean view of evidence"—"the attempted explication of epistemological relations in terms of causal relations." † Joseph attacks this foundation, in particular Quine's contention that a "realistic theory of evidence must be inseparable from the psychology of stimulus and response, as applied to sentences."¹ For Quine the

¹³ For instance: How does the fact that an action is required by the code whose promulgation or support would maximize utility warrant the conclusion that such an action is obligatory? The persuasiveness of rule-utilitarianism as an account of right-making characteristics is in one sense weakened substantially, I think, by the move away from the grounding of rightness in the consequences of the ideal situation in which the code is accepted.

If we have good reasons to obey the provisions of such a code, it may well be that this is *not* because to do so is obligatory, but rather because to do so will likely aid in its promulgation.

* This paper was read before the 1978 meetings of the Southern Society for Philosophy and Psychology.

† In "Conventionalism and Physical Holism," this JOURNAL, LXXIV, 8 (August 1977): 439-462. Pages follow citations in parentheses.

¹ *Word and Object* (Cambridge, Mass.: MIT Press, 1960), p. 17; cited by Joseph, p. 458.