IDEAL MORAL CODES

Duncan MacIntosh
Dalhousie University

I. A Problem for Ideal Code Theories

Standardly, a moral code is said to be the best one just if accepting it would maximize the existence of some morally desirable state of affairs. That state of affairs could feature anything from a level of aggregate utility, to people's negative rights being respected, to God's will being done, depending on the kind of moral theory. (I shall be commenting here on Ideal Rule Utilitarianism mostly—hereafter IRU—but my comments could apply as well to other forms of ideal moral code theory.) Gregory Trianosky[1] tells us that in IRU,

(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.'[2]

Following Richard Brandt, Trianosky takes acceptance of a code to involve being motivated to conform to it, being disposed to feel guilty if one knowingly transgresses it, and feeling disapproval towards those who violate it.[3] As I read it, this specifies a relation between acceptance of a code and compliance with it. If one accepts a code, one experiences no psychological barriers to compliance, and will comply if circumstances permit. The Brandtian characterization of acceptance is disputable, but no matter how minimal we make what is involved in accepting a code, the following problem can arise: It might in fact be psychologically impossible for all persons to whom the code is to apply to fully accept the code most ideal by this standard. Could it still be the best moral code?

Trianosky thinks this impossibility counts as some sort of objection against a code's being best, as in good slippery slope arguments. For example, if approval of euthanasia tends to weaken disapproval of unvarnished murder, then a code

---

Duncan MacIntosh (Ph.D. Toronto 1986) is an Associate Professor of Philosophy at Dalhousie University in Halifax, Nova Scotia, Canada. He has published on philosophy of language and science, epistemology, metaethics and decision theory, in Analysis, Canadian Journal of Philosophy *(forthcoming),* Dialogue, The Southern Journal of Philosophy, *and elsewhere.*

proscribing murder, period, would be superior to one proscribing murder and endorsing euthanasia, since attempting to accept the second code will cause more murder.[4]

Unfortunately, he points out, according to (1), such objections are irrelevant, for what there makes a code ideal is that it would maximize utility *if* universally fully accepted, whether or not the code happens to be psychologically accept*able* in actuality; the goodness of a code is measured by the consequences of the sometimes contrary-to-fact condition of its having been accepted.[5] How then, he asks, should IRU be modified to sensitize it to this sort of objection?

II. Two Possible Solutions

Trianosky reports John Rawls as suggesting that only codes which it is psychologically possible to accept can be ideal.[6] This would make slippery slope objections both relevant and decisive. L. A. Whitt concurs.[7] The criteria of the goodness of a moral code ought to be modified so that only psychologically acceptable codes are even starters. She thinks codes which have best consequences only in the event of acceptance by psychologically *ideal* agents (ones who could accept codes inacceptable for us) are too 'otherworldly' for actual agents, and are to be ruled out on grounds of being irrelevant for them. Thus:

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

But (2) eliminates codes the *attempted reinforcing* of which would maximize utility over the attempted reinforcing of any other code, and Trianosky thinks that such codes ought to be able to count as best. To take a case (not his), it might be impossible for us to respect others' rights *all* the time. But attempting to reinforce a code which always required that respect might produce more right-respecting than otherwise, making the code a good one. To protect such codes, he thinks that the criterion of code goodness needs to be amended to weaken or supplant the 'success' condition; thus a code is not best just if its *successful* acceptance would maximize utility, but just if its *attempted reinforcement* or *promulgation* would do so. Promulgation involves 'the systematic utilization of certain psychological tools for the purpose of inculcating and maintaining acceptance, here without the assumption that such efforts would necessarily [universally and fully] succeed.'[8] For him then, the psychological inacceptability of a code is only decisive where it so diminishes the utility

390

accruing to promulgation of the code that its promulgation would not maximize utility compared to promulgating any other code. Thus,

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

In this paper, I find several difficulties with (3). I argue that it is an incomplete solution to the original problem, while enhancements to it in its original spirit would do violence to the idea that the right code would give a correct standard of behavior. It also estranges our moral obligations from the dictates of a correct moral code and engenders unacceptable moral paradoxes. I then discuss the possibility that the act-utilitarianism implicit in Trianosky and the rule-utilitarianism implicit in Rawls/Whitt may be incommensurable. But I conclude that the virtues of the two forms of utilitarianism can be combined in a new standard for the goodness of a moral code: A code is the best one just if more utility results the closer we come to full and universal acceptance of and compliance with it compared to doing the same with any other code. We are then obliged to come as close as possible to accepting and complying with that code. This proposal avoids the objections to and combines the virtues of (1)-(3), while preserving the best features of act and rule utilitarian thinking about moral codes. It concedes to the act utilitarian that actions influenced by a truly good code must be utility-conducive, while vindicating the rule utilitarian thesis that our obligations to the code are related to the fact that its acceptance would maximize utility; the best code must be good both in principle and in practice.

III. Objections to Trianosky's Proposal

To begin then, I do not think (3) will sensitize IRU to all forms of slippery slope objection from psychological impossibility. For suppose there is a code which, if promulgated, would yield more utility than would accrue to promulgating any other code, but which it is in fact impossible for people to promulgate. Perhaps promulgating one part of the code makes another part of it impromulgatable. Or people may lack the requisite knowledge, or 'tools,' or may not be able to get themselves to embrace as a good thing, aiming at the acceptance of the code they are supposed to promulgate. Can it still be the most ideal code? Presumably we have insensitivity to the slippery slope back again, with no way to deal with it at this juncture in (3). Indeed, instead of dispensing with success conditions on code evaluation

altogether, (3) in fact merely contains a *different* success notion; rather than conditioning the goodness of a code to its utility upon *acceptance*, it is conditioned to its utility upon *promulgation*.

Is *promulgation* really a success notion? One might argue that promulgation stands to acceptance as trying stands to doing, only 'acceptance' and 'doing' constituting success notions proper. But one can either try or not, and surely if one cannot, the actual attempted reinforcement utility of the code is not available to justify it as the best code. One might doubt whether there could really be a code which one could not even try to promulgate. But consider the difference between, first, accepting and following a code which enjoins one to respect people's rights, second, trying to get oneself to accept such a code, and third, wanting to be able to at least try to accept such a code because one has heard that something called 'respecting people's rights' would be a wonderful thing to do, but just not having the concept of a person's rights, and so not knowing how to go about trying.

How might we treat this difficulty in the same spirit as Trianosky's original proposal? He wanted to eliminate the success condition because the maximization of utility might be achieved by the undertaking of an attitude or course of conduct respecting a code short of fully accepting it, and this virtue of a code would be ignored if we took the psychological impossibility of its full and universal acceptance to be decisive against it. Presumably then an impromulgatable code could still be ideal if taking some attitude to or action in respect of the code weaker still than that of promulgating it would maximize utility over taking any other act or attitude respecting any other code.

But take any combination of attitude and code; it is possible that the required attitude is psychologically impossible in some degree for some or all of the code for some or all of those to whom it is to apply. Could it still be the best code?

We might always weaken the attitude requirement on the same rationale as before, but the problem can always rearise in principle. And I think there are limits to the viability of this strategy, ones showing that ultimately, the criteria of goodness for moral codes *must* contain a success notion of some sort, and further, in keeping with the Rawls/Whitt proposal, one that must be satisfiable for a code to be a 'starter.' For there are attitudes to codes the taking of which might maximize utility, but such that it would be logically bizarre to suppose that this argued the virtuousness of a code as a guide to morally correct behavior.

392

It is certainly possible, for example, that maximum utility would accrue to promulgating *not* following a certain code. Promulgating non-compliance is the taking of a less strong pro-attitude towards a code than that of promulgating compliance. Indeed, as the negative limit of trying to follow a code, it is, perhaps, on the continuum of attitudes to codes which Trianosky's strategy (if not Trainosky himself) could countenance. But it approaches absurdity to think that this code/attitude complex can be endorsed on putatively rule utilitarian grounds. Can a code really be the *best* one precisely *because* reinforcing *not* following it maximizes utility?

Weakening the attitude requirement leaves open yet another possible condition under which utility would be maximized, vis., where maximal utility would occur where no particular attitude is taken to the code at all, as in cases of indifference to, or ignorance of, a code. Ignorance of the code of conduct associated with the *Dungeons and Dragons* game cult, for example, is actually utility conducive for high-school students, some of whom, if they do learn of it, will be highly susceptible to its inutile suasions, and will be led by it to make and implement suicide pacts.

Both the extremes of trying not to follow a code and of being indifferent to or ignorant of it are possible conditions under which maximum utility might issue, and yet are surely 'falsehood conditions,' if you will, of rule utilitarianism. Taking a negative attitude towards a code is the very paradigm of not following it at all. While maximum utility resulting from taking no express attitude towards a code implies that there is no rule to which a rule utilitarian justification applies on these assumptions. In both cases, the rule utilitarian justification for the goodness of a code fails—in one case, because what is justified is precisely *not* implementing a given code, and in the other, because no *express* attitude whatever to any code is justified.

The logical oddness of counting these factors as contributory to a code's moral correctness is even starker in the following case. Suppose there is a code which, if printed up, would be ideal in weight for use as a throwing weapon. Other codes are too heavy to throw, yet others too light to do any damage. It might be that this code is thus the trustiest one to carry in one's pocket while navigating the streets; a handy defense against ruffians, just right for knocking incautious gavagi out of the way of speeding cars, etc. Liking the code for its projectile utility is the taking of a positive attitude towards the code, and being disposed to throw the code has high utility. But it would be crazy to say that this in any way makes it

the right guide to morally correct behavior. (One might object that this and the preceding cases are indeed crazy—too crazy to show anything. But what they show is that utility cannot accrue to a code through just any old process and then count towards the moral correctness of the rules in the code. And the farther we get from the code's utility resulting from compliance with the code, the farther we get from it making sense to say the rules of the code are morally correct on grounds of its effects on utility.)

Further, though the utilitarian has always been embarrassed by not being able to justify selecting between *prima facie* moral and *prima facie* immoral methods to the maximization of utility in the same degrees, matters are now still worse. For there can be more than one way to generate a behavior (and the consequent utilities which render a code ideal) concordant with a given code of conduct. Imagine two codes, Never Steal, and Always Steal. Trying to *follow* Never Steal and trying *not* to follow Always Steal might well generate the same behaviors, and the same levels of utility. Which code should get the nod as the best code given that the resulting utilities would be the same either way? One might argue that trying to follow Never Steal just *is* the same as trying *not* to follow Always Steal—that it is logically impossible to do one without doing the other—so that there is really no choice, nor really any problem. But surely the world could be different depending on which method one used to secure these behaviors; the law books would be different, for instance, as would be the attitudes to them one would try to reinforce. In one case, one would do as the books say, respect the law, etc.; in the other, not. Yet there might be no difference in the utilities consequent upon each strategy. Now normally, you break ties by flipping a coin. But here, there is something wrong with finding it morally indifferent which we pick, for it seems bizarre to say that the *best* law could be the one which people should always try to *break*. (I do not, with this argument at least, dispute that the best law to have might be one that one might also be obliged, sometimes, to break. Good positive law might not always exhaust or accord with actual obligation. I only dispute that a law could be the best law *qua* law *because* breaking it has best consequences.)

Trianosky's stratagem may be too liberal in yet another respect. Imagine two codes, Give One-Third of Your Income in Taxes, and Give Nine-Tenths of Your Income in Taxes. Suppose it is psychologically possible to accept and to follow Give One-Third, but only to promulgate Give Nine-Tenths. Suppose accepting Give One-Third produces the same utility

as promulgating Give Nine-Tenths, and that there is no other code/attitude complex with a consequent utility even as high as that associated with these two codes. Which is the best code? Surely on the classical formulation of IRU, and granting that ought implies can, it is Give One-Third. We are obliged to follow the code with the highest compliance utility if it is followable, and cannot be obliged to follow a code we cannot follow, even if *attempting* to follow it has best consequences. But if we accept the spirit of Trianosky's liberation, we now have no grounds for morally preferring the one code to the other. There is no rationale for holding the code one could accept to be more ideal or obligating than the one which one could not accept. Put another way, where utility is not at issue, there is no way to give *any* weight to the ought-implies-can rule.

Now, I do not know whether a utilitarian takes himself to be obliged to generate a utilitarian foundation for every feature of morality, or whether he will allow some of them to be utility-independent (like the conditions under which people have the excuse of inability to do otherwise for conduct *prima facie* immoral). Perhaps, in defending this formulation of IRU, Trianosky would accept it as analytic that ought implies can. But if he admits the principle as an *a priori* constraint on correct conceptions of morality, codes which it is impossible to accept will be ruled out of court straight off, and Trianosky's proposal must yield to Rawls'. For then a code must be actually psychologically acceptable to be ideal; it cannot be correct if its being correct would oblige us to do the impossible (e.g., accept and comply with a psychologically inacceptable code). On the other hand, if, to escape this difficulty, he *repudiates* the ought-implies-can principle, I think he will have a most implausible conception of the extent of moral obligation. For how can we be obliged, on any rationale, to do something we just cannot do? (I consider below, ways for one to have an obligation with which one cannot comply; but I think there must then be a way to avoid incurring moral culpability for non-compliance.) While if to escape this Trianosky proposed that it follows from a code's being the best one not that acceptance of and compliance with it is obligatory, but rather that mere promulgation of it is obligatory, he will not have successfully defended IRU as an account of the right-making properties of actions. IRU, recall, is virtually *defined* as the view that the right actions are those prescribed in the right code.

But perhaps Trianosky would say that his account *does* preserve that feature. It is just that the code which one is

obliged to accept and comply with on his account, is a kind of 'shadow' code to the one one is obliged to promulgate: one must accept a code prescribing promulgation of the code promulgation of which maximizes. But then he would in fact have the classical account of IRU—the right code is the one acceptance of which maximizes, and one must comply with the right code—though the classical rule utilitarian might be surprised at the content of the codes he is obliged to accept. They do not say 'do x,' but 'promulgate doing x.' But again, what if one *cannot* accept the shadow code? We are back where we started.

There is a further difficulty with Trianosky's proposal deriving from the fact that the utility of a code upon its mere promulgation can diverge from its utility upon actual universal full acceptance. We have not far to seek for an example of a code which it maximizes to merely promulgate (i.e., to unsuccessfully reinforce), but which it does not maximize to actually accept and comply with. For as Gregory Kavka has pointed out,[9] the consequences of public and private attitudes towards actions can be good, while the consequences of the actions themselves can be bad. To adapt his infamous case to our problem, it might be maximizing to promulgate a code demanding retaliation upon nuclear attack, for the deterrent effect of so promulgating. Yet it might not be maximizing to fully accept and comply with that code, for that will result in a pointless and inutile retaliation in the event of a cataclysmic attack. (Note: in Kavka's original example, only full acceptance of the retaliation code would tend to deter; but in my adaptation, promulgation of the code alone is stipulated to have this effect.)

Now one might think that (3) is unaffected by this, for two reasons. First, by (3), what makes a code right is that promulgating it maximizes. Thus its acceptance utility is irrelevant; if accepting a code is inutile, that is no objection to its virtuousness by the promulgation utility standard. Second, since acceptance of the code is by hypothesis impossible, perhaps we need not be concerned with the bad consequences of actually accepting the code it is maximizing to merely promulgate.

However, I think Trianosky must be disturbed by the example, again, for two reasons. First, he thinks that accepting and complying with a code is right just in case doing so promulgates the acceptance of the code. For he thinks a code is made best by its promulgation having the effect of maximizing utility, that one's obligation to the code is to promulgate it, and that accepting and complying with a code

396

can often be the most efficacious instrument to its promulgation.[10] Let us assume acceptance promulgates the code in question. (Perhaps the Commander-in-Chief's willingness to retaliate induces those in his command to perform their pre-attack duties under the retaliation code more efficiently.) We are now obliged to accept the code if we can, not just to verbally promulgate it. But by hypothesis, accepting the code would be inutile. And Trianosky's aim in offering (3) is to capitalize on opportunities for maximizing utility. Thus, one is apparently obliged by his (3) to accept the code with submaximal acceptance utility because doing so is instrumental to promulgating the code, but is obliged by his aim in offering the proposal not to comply, to avoid disutility. His (3) then, is incompatible with his aim in offering it when there are conflicts between the promulgation and the acceptance/compliance utility of codes. One is obliged to promulgate, for the great consequences of *merely* promulgating, obliged not to do what is necessary to promulgate, for the terrible consequences of *succeeding* at inducing acceptance and compliance.

If we consider just Trianosky's aim of rewriting ideal code theory to take advantage of opportunities to maximize utility, and consider his strategy of separating the actions prescribed in the right code from the comportments we are obliged to undertake to the code, we run into a classical puzzle. When the promulgation and acceptance utilities of a code diverge, we have such paradoxes as that it is best (because maximizing) to intend to follow, desire to follow, reinforce the following of (and so on down the gamut of positive attitudes to the code), a code which it would be wrong to actually accept and comply with (because that would be non-maximizing). It seems then that Trianosky's proposal inherits Kavka's paradox that it can be rationally and morally obligatory to intend (etc.) an irrational and immoral action. And I take this possibility to be a *reductio ad absurdum* of any moral theory in which it is a possibility. For it violates yet another constraint on the rightness of a code, vis., that a code is wrong if it prescribes wrong actions, a constraint comparable in force and intuitiveness to the ought-implies-can principle. The lesson to be got from Kavka's cases is that no consistent morality can evaluate actions both by the consequences of their performance independently of the consequences of public and private attitudes towards their performance, and by the latter independently of the former, for their consequences can differ in utility.

IV. A New Solution

Returning now to the larger conflict between Whitt/Rawls and Trianosky, we appear to be left with a vicious dilemma. The principle which animates Trianosky's proposal is too liberal, letting in obligations we do not in fact have (because we *cannot* fulfill them), and counting codes as best when the way they contribute to utility seems almost incidental to what they prescribe, indeed, perverse relative to their prescriptions (as when reinforcing noncompliance with a code maximizes). And the principle may, indeed, be finally self-conflicting, or at least in conflict with the motives for advocating it (in the case where mere promulgation maximizes, but successful compliance does not). Moreover, it does not, consistently with IRU, make the right actions identical to those prescribed in the right codes. (The promulgation utility of a code makes it right to promulgate it, but not necessarily to comply with it.) But the Whitt/Rawls alternative is too strict, for to opt for it is to loose opportunities to maximize utility by promulgating inacceptable codes, since we cannot count a code ideal by their standard if we cannot accept it.

We thus need a criterion of the goodness of a moral code with the following properties: It must incorporate a conception of obligatory attitudes to a code such that taking them to the code makes it responsible for the resulting utilities in the right way, i.e., in a way tracing from its properties *qua* (heeded, in some sense of 'heeded') code, for its illocutionary status as giving a positive standard for behavior. It must make certain slippery slope objections to a code's goodness relevant but not necessarily decisive. It must be compatible with the analytic ought-implies-can constraint on possible obligation. Finally, it must not paradoxically oblige one to take some kind of pro-attitude to an action one is obliged not to actually perform, nor to perform an action to which one is obliged not to take any kind of pro-attitude. Moreover, we would be wise to take account of the important insight in Trianosky: Utility is not just a consequence of a code, but of a combination of code, attitude to the code, and circumstance. The objective here is to give a recipe for a code/attitude combination that exploits circumstance to yield utility, and to give a conception of the relation between this recipe and our obligations. The two should conspire to explain how the resultant utility relates to such obligations as there may be to comply with the prescriptions of the code.

A proposal: We should take a code to be the best one just if the following conditions are satisfied. (a) Were it to be

398

accepted (where acceptance normally implies compliance), utility would be maximized over acceptance of any other code. (b) The taking of some weaker but nonetheless positive attitude to the code *qua* positive behavioral standard, would maximize over accepting any other acceptable code, over taking any comparable weaker pro-attitude to any other code, and over taking any attitude weaker still to the code presently in question or to any other code. Finally, (c) there must be at least one pro-attitude to the code one can in fact take which satisfies (b). That is what makes a code the best code. But how does its being the best code connect up with our obligations and explain the rightness of various actions? Just in this way: People are obliged to come as close to accepting and complying with the best code as they actually can, and must come in some positive degree close. Thus,

(4) An act is right if and only if it expresses an attitude as close as is actually psychologically possible to that of acceptance of the code acceptance of which in the agent's society would maximize utility (assuming acceptance normally entails compliance, and assuming approximations to acceptance normally entail approximations to compliance), where so approximating acceptance maximizes utility over comparably so approximating acceptance of any other code.

Now, (4) specifies the right act derivatively from the right attitudes, which are in turn specified derivatively from the right code, which is identified by its comparative utility on taking pro-attitudes to it up to and including those of universal full acceptance and compliance. Roughly speaking, one is obliged to try as hard as one can to do what it would be ideal for one to do, namely, what would actually be compliant with rules universal full acceptance of and compliance with which would maximize. But there is a proviso on one having this obligation: Ever-closer approximations to universal full acceptance and compliance must have ever-better consequences, and no other psychologically possible attitudes to this or any other code may have better ones.

Some of the terms and concepts I have just introduced need clarification. First, what is it for an action to express an attitude to a code *qua* positive standard of behavior? Well, for example, full acceptance of the code would be expressed in perfect compliance with the code (i.e., always doing everything it requires), circumstances permitting; it just consists in internalizing the code's precepts as one's principles of action. Full rejection would be expressed in complete non-compliance; it is to internalize principles of action opposed to the code's. In between full acceptance and full rejection, we have such things as full acceptance of part of the code,

399

rejection of other parts. Here, we can expect perfect compliance with part of the code, non-compliance with other parts. We can also envisage weak acceptance of some or all of the code. This is the acquiring of motives for action sufficient to produce full compliance with parts of the code, or to produce partial compliance with some or all of it. Obviously the details will depend on the nature of the code and the attitudes taken in respect of it. Thus, suppose a code demands that one give 70% of one's income to charity. Perhaps one can only acquire motives of generosity sufficient to result in one giving 50%. Acquiring those motives would consist in partially (in the sense of weakly) accepting the code, giving 50% consisting in partial compliance with it. Here, approximations to acceptance consist in various commitments to donating some percentage of one's income, approximations to compliance consisting in actually donating some percentage. Or suppose a code demands that one never lie. Partial acceptance might consist in acquiring motives sufficient to result in one not lying 60% of the time, or in never lying in certain circumstances, sometimes lying in others, etc. If we imagine there being some distribution of such motives and actions among the agents of some society, we see what it would be like for there to be approximations to universal full acceptance and compliance in a whole society.[11]

We have some idea, then, of what partial approximation to full acceptance and compliance might involve for a code, and can make some sense of different degrees of approximation along dimensions determined by the code's character. The more difficult concept is that of equal degrees of acceptance/compliance for different codes. It seems we must make sense of this idea if we are to choose between codes by the best one having a higher utility at every degree of acceptance and compliance compared with all other codes at equivalent degrees. But what counts as an equivalent degree of acceptance or compliance for different codes? They can vary widely in content and structure. One might have but one clause, another, ten clauses. Is perfect acceptance of five of the latter clauses equal in degree of acceptance to half-hearted acceptance of the former code? Is there always even a fact of the matter about this? Does the very idea of a unified system of ranking and scoring codes make sense? Such worries parallel those which arise in explaining what it is for scientific theories to approximate to the truth, and for possible worlds to most closely resemble the actual world, which notion may have to make sense if we are to account for how the counter-factual claims of science can have determinate truth values.

400

Is there really a single dimension of truth-likeness along which we can arrange apparently incommensurable theories(?); a single dimension of resemblance along which we can order wildly different possible worlds by degree of closeness(?); a single dimension of degrees of acceptance and compliance along which we can arrange attitudes towards and actions respecting structurally dissimilar moral codes?

There has been much progress in explicating the concept of approximation in the cognate fields just mentioned, giving hope that something similar could be done for ethics. Moreover, while I have nothing new to offer on this (nor space to do so here anyway), and while I acknowledge the above difficulties, I do think that we may already understand the concept well enough to use it in at least rough and ready ways in all cases, and even in precise ways in some cases.[12] But the present enterprise may be viable even if the concept of equivalent approximations systematically resists unified explication. For we can clearly make rough sense, for any given code, of degrees of acceptance of and compliance with it *in its own terms* (even if many different comportments regarding it will count as equal in degree of acceptance and compliance, and even if some comportments do not easily rank). And we can speculate on the utility consequent upon each level of acceptance and compliance for each code. Thus we can already make the comparisons between codes I require. For any two codes, A and B, A is the best code (compared with B) just when all of the following conditions are satisfied. (i) The actual utility resulting from the maximum actual degree of acceptance/compliance currently possible for A (given current human nature and resources, etc.) is higher than that imagined for the maximum actual degree of acceptance/ compliance currently possible for B. (Note that the two degrees need not be commensurable; we need only the best actually possible level of acceptance and compliance for each code *in its own terms*.) (ii) The utility from any level of acceptance/ compliance for A and B weaker (in their respective terms) than the actual maximum level possible for A and B is lower than the maximum actually possible for A. (iii) Conditions (i) and (ii) are still best met by A no matter what degrees of acceptance/compliance we stipulate to be the maximum actually attainable levels for the two codes in their own terms in current circumstances. (iv) The utility from hypothetical perfect acceptance/compliance for A is higher than that for the same for B.

These conditions explicate (a)-(b), above. Thus clause (iv) guarantees that A is maximizing on perfect acceptance/

401

compliance. Clause (i) guarantees that A is maximizing on current best approximations to perfect acceptance/compliance. Clause (ii) guarantees that the current best approximation to perfect acceptance and compliance for A is maximizing compared to any weaker currently possible approximations for A or any other code. And clause (iii) guarantees that under any common conditions whatever, the approximations there possible for A and B are such that the utility consequent upon that approximation for A is higher than that for B.

This avoids us having to specify a common measure of approximation for what may be structurally dissimilar codes. For it derives the superior virtuousness of a code from the idea that under any common conditions whatever, where these conditions are specified *independently* of the idea of common levels of approximation for codes, the best code will be one such that the level of approximation (in its own terms) possible for it under those conditions will result in more utility than the level of approximation (in its own terms) possible for any other code under similar conditions. And we need not rank conditions by the levels of approximation they permit for different codes in some ranking in common between codes; we need only compare codes by the utilities generated by possible approximations in each code's terms for each set of conditions. For if one code beats all others on each possible condition, then it would still do so no matter how those conditions were ordered on the scale of permitting common degrees of approximation for different codes. Thus we can let this fall where it may and still compare codes at each degree of compliance, however ordered.

Note that (4) is not itself a code, and its decription of one's obligations does not exist as a code. An analogy: The maximizing conception of rationality gives the principles of rational choice (vis., to be rational is to maximize one's individual expected utility), but a rational chooser does not try tó conform to the maximizing principle. He simply tries to get what he wants. Conforming to the maximizing principle is not his aim; rather, the principle describes how he will have chosen if his desires justify his actions. Similarly, (4) describes how a moral agent will choose if his action is justified by the morally correct code.

There are, I think, several virtues in (4). First, one can never be obliged to take a pro-attitude to an act it would not be right to have performed, since one is expressly required to take pro-attitudes only to acts it would be maximizing for all to perform. Second, one is not obliged to do anything one

402

cannot in fact do, since one need only come as close as one can to acceptance of and compliance with the ideal code, and that is consistent with perfect acceptance/compliance being impossible. Third, it is relevant to a code's goodness whether taking pro attitudes to it is actually possible, since if one cannot take any pro attitude to it, it is discounted from being an ideal code. And it is relevant to one's obligations to conform to the code whether one can in fact obey it, for if one cannot even approximate compliance with it, one has no obligations whatever with respect to it. Fourth, the best code is one the good effects of the existence of which *qua* positive standard of behavior are such that it actually makes logical sense to count them towards the moral goodness or idealness of the code *qua* code (rather than, say, *qua* projectile). For what counts towards its goodness is just that the closer everyone comes to fully accepting and complying with it, the better the consequences. And the *way* the code has its effects is that the closer people come to accepting and complying with it, the better are the consequences. Fifth, what makes the code the morally correct one is something actually relevant to such obligations as there may be to actually conform to it (for such obligations arise just when one *can* conform to it), and what makes one have an obligation to approximate acceptance of and conformity to it is just that ever closer approximations have ever better consequences.

How well does this proposal deal with the original counter-examples? It rules out codes which proscribe murder but prescribe euthanasia, assuming a code proscribing murder simpliciter has a higher acceptance utility than the approximation utility of the do not murder/do perform euthanasia code. For the latter violates (b). (Of course this does not automatically endorse the former.) But it rules in the code which says to always respect people's rights, for, *ex hypothesi*, its acceptance/compliance utility is maximal—so that it satisfies (a)—and its approximation utility allows it to meet (b); better to respect people's rights sometimes or often rather than never. Moreover, since, by hypothesis, it is possible to respect people's rights some of the time, it meets (c). Finally, if it would not be maximizing to reinforce retaliation, to commit to retaliation, and to retaliate, then the code saying to retaliate cannot be the ideal code.[13]

V. Some Possible Objections

Still, we may have doubts about (4). For one thing, it seems only to oblige us to come as close as we can to acting in an ideal way, not necessarily to actually act in an ideal way.

But are not our obligations to *comply* with an ideal, not merely to *approximate* compliance? It is not enough that one merely try to refrain from murder, for example; one must actually refrain.

On this, two observations: First, remember, on (4), the right code does not merely say 'try to do x,' but 'do x.' But as your ability to comply decreases, so does your culpability for non-compliance. This is consistent with one's obligations being to comply with ideals of conduct, and with one's incapacity to comply not defeating those obligations. For to say that one is obliged to comply with an ideal is just to say that one is morally culpable for deviations from it *except* given things like excuses—"I could not help it; I killed on instinct when he attacked"—or explanations—"I did not know the drug would kill him; I was told it would save him"—or justifications—"I had to kill him to save the others." One fails to act as obliged, but has not acted immorally. One's action is not morally right by the standard of the ideal code, but neither is one morally culpable for one's deviations under those conditions. So to approximate compliance to an ideal of conduct as best one can is to approximate to the point at which excuses, explanations, and justifications kick in. Presumably the ideal code, or the ancillary apparatus for its application and enforcement, would contain provision for such ameliorations. The thing is, the code still applies because one needs an excuse, explanation, or justification for not doing what it says; but because one has one of these, one is not culpable for one's non-compliance.

But if we do not hold people morally culpable for *all* deviations from the ideal, will they not be less well motivated to conform to it? Perhaps, but it can be equally demotivating when a code embodies unreasonable requirements; it is then disrespected for being unrealistic or unfair. A reply to both worries: the right code and the right set of ameliorators component in or ancillary to the code, would be as demanding as it maximizes for them to be. Thus, if encoding something, or if treating something as an obligatory level of approximation to the code, causes increases in the extent and perfection of approximation or compliance up to the level the code demands, and if such increases in turn are ever-more-utility-productive, then they should be encoded. But demands stronger than that, ones which, while they increase the frequency of a presumptively desirable behavior, explicitly require the impossible of people, must not be encoded unameliorated, owing to the ought-implies can principle. No

404

correct ethic can violate that principle, so we should interpret our obligations to a code to reflect it.

While we have just worried that (4)'s ideal codes will not be sufficiently morally demanding on our conduct, one might also object that (4) sets impossible standards for something's being the morally right code. For it is possible, perhaps even likely, that there *is* no moral code with both the best actually possible approximation utility and the highest full acceptance/compliance utility.

Nonetheless, I stand firm. If no code is both approximation and acceptance/compliance maximal compared to other codes, then either IRU is a false theory of moral obligation (for even in what I claim is its most plausible version, it fails to correctly identify any obligations); or, while it is the true theory—while it does capture what would be true of moral obligations if there are any—there are in fact no moral obligations.

For consider the alternative. Suppose code A is approximation maximal—no other code is such that the maximum actually possible approximations to it result in a higher utility than that resulting from the maximum actually possible approximation to A. But suppose that were another code, B, to be accepted, more utility would result than from the acceptance of A. A is approximation maximal, B acceptance maximal. It would be disastrous to accept A, disastrous to quixotically reinforce compliance with B. A cannot encode our obligations because it would be non-maximizing and so wrong to succeed in complying with it. B cannot encode our obligations, because since all our attempts to reinforce compliance with it must fail, and such failures are inutile, its manifestation in our society will cause more harm than good. To say that A encodes our obligations would be to say it was right to aim at a wrong action. To say that B encodes our obligations would be to say that it was wrong to aim at a right action. Thus, here, there is *no* code now that stands as the right guide to morally correct behavior. We may be obliged to (merely) attempt compliance with A for the good consequences of so attempting, but that obligation is not in any code. Were conditions different, so that we could accept the other code, then it might be right to comply with it. But since conditions are as they are, we have no obligation to comply with that code. Thus, none of our present obligations figure in a code; here, IRU is false. Of course, there may come to be an ideal code for us if what is psychologically possible for us changes so that we can then accept and/or approximate acceptance of, a code we could not approximate or accept before, a code which would then be both approximation and

acceptance maximal. Moreover, we may have obligations on, say, act utilitarian grounds, to attempt to alter our psychologies so that some code will become right for us by (4); it may be maximizing for us to alter ourselves so that we can ever more perfectly approximate a code it would be maximizing to approximate ever closer up to perfect acceptance and compliance. But as of now, our obligations are not found in any code.

VI. Conclusion

Proposal (4), then, seems to satisfy all of the desiderata we uncovered in trying to reconcile Rawls' and Trianosky's objectives. Unlike classical IRU,what makes an action right by (4) is not simply that it conforms to a code universal full acceptance of which would maximize, but that it comes as close as possible to acceptance of (and compliance with) a code which has ever better consequences compared with any other code the closer everyone comes to accepting (and complying with) it. Unlike the Rawls/Whitt proposal, by (4) something can be the right code even if it is psychologically impossible for it to be fully and universally accepted; it merely needs to be possible for it to be approximated in some degree. And unlike Trianosky's proposal, first, (4) guarantees the consistency of moral attitudes with moral actions (since actions are right only if they express right attitudes, namely, maximizing attitudes to maximizing actions); second, every obligation is directly justified by the utility of meeting it; third, the goodness of actions is measured by their positive degree of approximation to codes maximizing on acceptance (among other conditions). The resulting theory looks more rule utilitarian than Trianosky's, and yet, unlike the Rawls/Whitt proposal, it is sufficiently flexible not to rule out a code simply on grounds of its being unacceptable. But since there must be *some* appropriate psychologically possible pro-attitude towards the code, slippery slope arguments could show that certain codes do not satisfy *that* requirement. This effectively places psychological limitations on which codes could be ideal. Nevertheless, a code can be the best one even if its full and universal acceptance is impossible.

It may have seemed, initially, that there were two kinds of virtues a code could have: it could be good on acceptance, or good on being reinforced. It may have seemed that the former are the codes we should follow if we were ideal beings, perhaps the codes we should try to modify ourselves to be able to follow; and that the latter codes are the ones the following of which we should reinforce given that we are

406

imperfect. But I have suggested that neither code captures our current obligations, and that neither of those virtues alone makes a code the right guide to currently morally correct behavior. The best code must have ever better consequences the closer we all come to fully accepting and complying with it compared with similar comportments to any other code. We are obliged to come as close as we actually can to acceptance of and compliance with that code, if our obligations are indeed to be found in any code at all.[14]

NOTES

[1] Gregory Trainosky, 'Rule-Utilitarianism and the Slippery Slope,' *The Journal of Philosophy*, 75 (1978), pp. 414-424.

[2] *Ibid.*, p. 416.

[3] *Ibid.*, p. 416.

[4] *Ibid.*, p. 418.

[5] *Ibid.*, pp. 418-419.

[6] *Ibid.*, p. 422.

[7] L. A. Whitt, 'Acceptance and the Problem of Slippery-Slope Insensitivity in Rule-Utilitarianism,' *Dialogue*, XXIII (1984), pp. 649-659.

[8] Trainosky, 'Rule-Utilitarianism,' p. 424.

[9] Gregory Kavka, 'Some Paradoxes of Deterrence,' *The Journal of Philosophy*, 75 (1978), pp. 285-302.

[10] This may seem odd. For is not his proposal that an action is obligatory if required by the code promulgation of which maximizes? And so is not the obligatoriness of an action a separate question from that of whether it promulgates the code? Further, cannot an action then be correct even if it does not promulgate the code (and even if it is an inutile action, provided it is an action compliant with the code promulgation of which maximizes)? (My thanks to an anonymous referee for this worry.) One might think so. Yet Trianosky says, "How does the fact that an action is required by the code whose promulgation . . . would maximize utility warrant the conclusion that such an action is obligatory? . . . 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 [directly? D. M.] obligatory, but rather because to do so will likely aid in its promulgation." (Trianosky, 'Rule-Utilitarianism,' p. 424, n. 13.) He seems to be assuming here that whatever comportment respecting a code is such that when everyone adopts that comportment to the code, the code is maximizing, is a comportment we are then obliged to adopt to the code (if the correct theory of what makes a code good involves *that* comportment). So if that comportment is universal compliance with the code, we are obliged to comply with it; if universal acceptance, we are obliged to accept it; if promulgation, we are obliged to promulgate it. In the first case, one is, perhaps, directly obliged to comply; in the second, one is, perhaps, obliged to comply because that is either a logical condition on, or a causal consequence of, acceptance (depending on how we understand acceptance); in the third, one is obliged to comply so far as that is the best way to promulgate.

[11] My thanks to an anonymous referee for the foregoing examples, and for encouraging a clarification of the concept of approximation.

[12] The problem pervades moral theory. For example, nearly every explication of morality uses the idea of people's intentions and actions coming more or less close to some ideal standard, and we sometimes speak of people being equally moral though they espouse different standards. This means we must

407

be able to compare agents on the degree to which they conform to various moral standards in the ways appropriate to each. The problem also figures in all the positions in the current debate. Indeed, that the concept of approximation to acceptance and compliance makes sense is a precondition of the debate. Trianosky, for example, speaks of comparing the utility of accepting one code with that of reinforcing another, reinforcement standing on the same continuum with acceptance and non-acceptance as an imperfect kind or degree of acceptance. Rawls and Whitt too imagine comparing the utilities of different codes under different degrees of acceptance and implementation; e.g., Whitt speaks of codes being broadly but not universally accepted. Since all of these theories use the concept, the possible difficulty of its explication is no reason to favor (1)-(3) over (4). In any case, the pervasiveness of the concept suggests that we have some intuitive and rough idea of what it is for codes to be equally completely or incompletely accepted and implemented.

[13] Note: I am neither repudiating euthanasia, nor endorsing a morality of rights, nor advocating a policy of non-deterrence. I am merely pointing out how (4) works in treating the foregoing cases on the attendant stipulated assumptions about the consequences of approximating and of achieving acceptance/compliance.

408