Kantian Ethics and Utopian Thinking

THOMAS E. HILL, JR.

§1. Problem cases: (A) Lying to the murderer and (B) Torture and the ticking bomb

WHEN, IF EVER, IS IT JUSTIFIED TO TELL A LIE? Bearing false witness is forbidden by scriptures and is subject to severe penalties in the judicial system. And yet most people seem to find it acceptable to tell lies in other contexts—to make a "surprise birthday party" possible, to protect someone's feelings (e.g. when asked 'Do I look fat?'), to protect one's cover when spying for one's country, etc. We generally agree that ideally, we would almost always deal with each other openly and honestly, but to be realistic we make exceptions. As we know, people draw the line about which lies are OK at different places. At West Point in the repugnant all–male days, I was told, the Honor Code was understood to forbid the cadets from ever telling a lie—except to a woman. Some have held that it is *always* wrong to tell a lie; for example, the philosopher Immanuel Kant (1797) famously defended the position that it is wrong even to tell a lie to a murderer in an effort to keep him from killing his intended victim who is hiding in your house. In the ideal world, of course, there will be no murderers and so no need to tell lies to protect their victims, but it seems foolish to infer from what we *would* do in that ideal world (where there are no murders) to what we *should* do the real world (where, sadly, there are murders).

Consider another sort of case. When, if ever, is it morally permissible to torture someone? For centuries torture was accepted as a legitimate practice, carried out by officials of both the church and the state, as authorized by official codes. In the 18th century in Europe there developed strong opposition, resulting in official bans on the practice that was supported by people with diverse philosophical and religious standpoints. The very word "torture" began to be understood as a negative moral label, like "murder," that *implies* that the act is wrong, not an acceptable means even to consider. And yet, as you know, some people defend the official practice of torture even today, and many more seem

Th. E. Hill, Jr. (✉)
University of North Carolina at Chapel Hill, USA
e-mail: thill@email.unc.edu

Disputatio. Philosophical Research Bulletin
Vol. 8, No. 11, Dec. 2019, pp. 505-529
ISSN: 2254-0601 | [EN] | **ARTICLE**

© The author(s) 2019. This work, published by Disputatio [www.disputatio.eu], is an Open Access article distributed under the terms of the *Creative Commons License* [BY–NC–ND]. The copy, distribution and public communication of this work will be according to the copyright notice (https://disputatio.eu/info/copyright/). For inquiries and permissions, please email: (✉) boletin@disputatio.eu.

to think that in extreme circumstances *secret* acts of torture (even if contrary to the law) are justified. The extreme case most often discussed is the "ticking bomb case" where we are to imagine that a *known* terrorist has planted a bomb to kill millions of people in NYC or Paris and he can *certainly* but *only* be stopped by the use of torture. The background conditions for this case are fanciful—absolutely certain knowledge that the torture is necessary and will be effective—but the problem is not restricted to these fictional cases. In an ideal world, presumably, there will be no terrorists and no need for torture, but once again, we must be cautious in inferring what we *should* do in the real world from what we believe we *would* do in a more perfect, ideal world.

These cases raise hard *practical* questions and also challenges to philosophers who propose general moral *theories*, but here I want to approach the problematic inferences from the ideal to the real world *indirectly* by first talking briefly about ideas of utopia and utopian thinking in political philosophy.

§2. What are "utopias" and "utopian thinking" (when meant as a criticism)?

From Plato to Rawls political philosophers have tried to expose our institutional defects, guide reform, and inspire hope by describing a more perfect social world. For example, in *The Republic*, Plato presented to the ancient Greeks his idea of a perfectly well–ordered and just city–state, guided and controlled by an ideally selected, trained, and tested ruling class. Details were attractive to some but repugnant to many: the ruling "guardians" were to have no private property, and no exclusive mating or marriage partners; children were to be held in common; poets were to be honored but driven out of the state; defective children were to be left to die; and the masses were to be told "the big lie" that their station in life was determined entirely by whether they were born "gold", "silver", or "bronze." But in his ideal city–state, there was social harmony, efficient division of labor, full employment, and gender equality that was radical for his time.

Centuries later, St. Thomas More (1516) described a fantasy island community that he called "*Utopia*" (a word that meant "no place"). Utopia was not presented as perfect in all respects (there were dangerous criminals, for example, who became slaves when caught). Nevertheless, More offered a picture of a world with social policies that might seem an improvement over those in Europe at the time—for example, religious toleration (though atheists were despised), full employment (though work was compulsory), lesser penalties for theft (though hanging was replaced by minor mutilation), and a somewhat more equal role for women (though they had to confess to their husbands every

month). More controversially, like Plato, More seemed to advocate devaluing wealth and private property. Prisoner's chains and chamber pots, for example, were to be made of gold in order to induce citizens to despise it; and citizens rotated houses every five years.

To pessimists, cynics, and many clear–sighted realists focusing attention on such **utopias** is futile and potentially dangerous. In fact, the term "**utopian thinking**" is now a pejorative label used to dismiss political arguments based on idealistic illusions. Marxists, for example, disparaged their bourgeois opponents by accusing them of utopian thinking[1] What counts as utopian and what value it has depend on many factors—the aims of the author, the audience addressed, the ways in which the projected ideals differ from things as they are, and their likely use and abuse by others.

Let us suppose for now a ***utopia*** is an idealized world or society presented by an author for practical purposes such as prompting particular reforms, proposing a goal to strive for, or inspiring hope that a better world is possible. "***Utopian thinking"*** now has a pejorative sense referring to the *misuse or abuse of a utopia*, especially the attempt to draw unwarranted recommendations from what would be appropriate in very different imaginary conditions. Utopian thinking of this bad sort, then, ignores relevant differences between ideal and real–world conditions, urging us to follow policies suitable for ideal conditions but impossible, impractical, dangerous, or otherwise inappropriate for us to try to institute or even approximate.

Plato's *Republic*, Thomas More's *Utopia* (1516), and John Rawls' *Political Liberalism* (1996) and *The Law of Peoples* (2001) are examples of *utopias*, I assume, but it is controversial whether the authors were guilty of *utopian thinking* in their political philosophies.

§3. Some potentially good uses of utopias

Legitimate uses of utopias arguably might serve these partially over–lapping purposes.

(A) **Suggest possible reforms to consider**. Utopias may *prompt critical reflection on accepted practices* and conventional wisdom, challenging common assumptions by posing even fictional alternatives that lead us

[1] Which, incidentally, was paradoxical given the Marxist's own utopian goals of a classless society and the withering away of the state (Engels 1878).

to wonder if there might be a better way to do things. As long as the utopia is not uncritically accepted and copied in disregard of relevant differences between fictional conditions and ours, the utopia may serve as a useful "stimulus package" for rethinking current ideas and practices.[2]

(B) **Vividly describe distant goals worth striving for.** Utopias may *inspire us to work towards a distant goal* that is worth striving for and may be possible to achieve. Even if we are uncertain whether an ideal can be achieved, we may have good reason to strive for it if coming close to achieving it would be of very great value and the evident costs and risks of trying are minimal by comparison.[3]

(C) **Present an inspiring ideal model to follow now.** A utopia might be offered and used, not as a distant goal, but as *a model for individuals to follow* or try to approximate in their conduct *now.*[4] Suppose, as many think, that there are some absolute principles and values we ought to adhere to in all circumstances, not just in a perfect world but in the real world and any circumstances we can realistically foresee. Commonly accepted examples include the principles that one should not enslave, rape, torture, or murder anyone. Many would add the prohibitions on child–parent incest, cannibalism, adultery, treating a person with contempt on account of race, gender, or religion. If principles such as these hold without exception, then a utopia might be constructed *as an inspiring picture* of how things might be if everyone conformed to these inflexible principles and values and *as a vivid model of how we should behave now* even though our world is corrupt and the ideal will never be achieved. Given the background presupposition that the principles of our utopia are valid without exception, using the utopia as a model for our behavior now would not be drawing illegitimate inferences from

[2] In *Utopia,* More (1516) evidently had this as an aim.

[3] In trying to present "realistic utopias" in *Political Liberalism* and *The Law of Peoples,* John Rawls (1996, 2001) apparently had this function of utopias in mind as at least part of his aim.

[4] This possible function of utopias occurred to me in thinking of Kant's briefly sketched "kingdom of ends" in the *Groundwork for the Metaphysics of Morals* (Kant 1785: Ak. 433–37, 233–36). This was not an "end to be produced" such that our moral task is to discover and take the most effective means to achieve it. Rather, it was both a sketch of a point of view for "legislating" moral "laws" and an idea of a morally desirable outcome that we can "think" is possible. If it serves to guide choices, the relevant principle is 'Follow the laws now that you and others *would* legislate as members of a kingdom of ends, whether others do so or not.' To make the ideal kingdom actual, Kant suggests, everyone would need to dutifully follow the its laws, but our duty to follow them does not depend on the likelihood of this happening.

the ideal world to the real world. We should follow the principles now as if we were in that utopia even though we know that in our real world there are rapists, murderers, torturers, racists, sexists, etc., who may never conform to the principles. Given the background assumptions, using the utopia as a model to follow now would be unobjectionable; but, of course, critics of utopian thinking will naturally question whether the inflexible values and principles built into the model are really, as supposed, valid and binding independently of context.

(D) **Inspire hope.** A utopia might serve *as a vision of a better possible world that inspires hope* regardless of whether it can legitimately serve to guide our decisions.[5] Because of factors beyond our control it may be that there is nothing we can do to bring about the better world, but still the description of a humanly possible better world might serve for some to inspire hope that can overcome despair and cynicism. A utopia might serve to inspire hope even when it cannot serve as a model for our conduct now[6] or as a realistic goal that we can hope to achieve by our efforts.

§4. Corresponding dangers of bad "utopian thinking" of all four types

Utopian thinking, or the use of utopias for *illegitimate* purposes, arguably comes in correspondingly many types. Let us review the potentially positive functions of using utopias that I have just discussed, and now look for the potential down–side

[5] This seems to be one of the functions of Kant's account of a better international order, *Toward Perpetual Peace,* especially in "First Supplement: On the guarantee of perpetual peace" (Kant 1795: Ak. 361–8, 85–92). The preliminary and definitive articles in that work were clearly meant to influence the actions of heads of state, but my suggestion is that the work was also meant to inspire hope beyond this action–guiding function. The message is not that rulers should make choices based on dreamy, soft–headed estimates of their chances of bringing about an ideal world, but rather that we should not despair when the outcomes look bleak but rather keep our morally grounded faith that, though largely out of our control, somehow eventually the natural forces behind wars, revolutions, and economic competition will accomplish what merely well–intended efforts cannot. I speculate that Rawls may have thought that his "realistic utopias" could serve a similar hope–inspiring purpose in addition to whatever action–guiding influence they have. Whatever its chance of being realized, his realistic utopias are supposed to exhibit a non–Hobbesian picture of a sustainably just world that human beings are at least capable of achieving. If so, returning to his early religious concerns, Rawls could see humanity as redeemable and moral struggles as not necessarily in vain. But this, again, is just speculation.

[6] For example, because its guiding principles are not valid independent of context as was assumed in the preceding case.

of each.

(A*) *Using a utopia uncritically as a guide to reform our practices* would be illegitimate if the particular practices that make sense in the ideal world context are useless, dangerous, or unsustainable in the real world. For example, in More's utopia no one may lock their doors, and in Plato's the rulers are always philosophers.

(B*) *Using a utopia as a goal to strive for* would be illegitimate if based on unwarranted optimism about its feasibility, underestimation of the costs of achieving it, and ignoring the risk that striving and failing for the ideal may be far worse than accepting "second best." (For example, even if Rousseau's ideal state (Rousseau 1762) would be better than ours, it would be foolish to scrap a more realistic system of checks and balances in futile efforts to establish Rousseau's ideal).

(C*) Using the principles of utopian world *as a model for how we should act in the real world* would be foolish, harmful, and morally indefensible if the validity of those principles depends on imagining that everyone always conforms to them. For example, being a pacifist makes sense in a world of pacifists, but it doesn't follow that in our violent world we should always refuse to fight.

(D*) Using a utopia *to inspire hope for a better world* can be foolish, delusional, a false comfort, and an obstacle to clear and effective thinking about immediate problems. Even if it does not mislead us into futile striving or following inappropriate principles, living with utopian *hopes* may make us too passive, out of touch with the here–and–now. For example, if I am always dreaming hopefully of glorious days that may somehow come in the future, on earth or in heaven, I may fail to take practical steps to make life better now. This, I suppose, is what Marx meant when he called other–worldly religion "the opiate of the masses" (Marx 1844: 71).

Those who propose and use utopias, I suppose, are not necessarily guilty of utopian thinking of these sorts, but these are always a danger.

§5. "Utopian thinking" in political philosophy: Rousseau's *Social Contract*

My main interest is in utopian thinking implicit in ordinary moral thought and in the moral theory of Immanuel Kant, broadly interpreted. As background, I find it helpful to look at the problems of utopianism in common sense morality and in Kant's moral theory as if the problems arise from reflecting on Jean–Jacques Rousseau's political philosophy. The broad "as if" story would run as follows. In the *Discourse on the Origin of Inequality* Rousseau traces the evils of his day to the corruption of basically good human dispositions by the conditions under which people became socialized (Rousseau 1755). Inequalities in wealth, power, and status bred envy, fear, greed, resentment, and civil disorder. The social contract that supposedly legitimized the property and power inequalities of the status quo was grossly unfair, a product of the "haves" duping the vulnerable "have–nots" to think that accepting their "chains" was for their own good. Later in *The Social Contract* Rousseau asked, in effect, what standard for laws *would* emerge from a fair contract made by free and equal people, not duped, pressured, or indoctrinated to accept an unjust status quo. The standard for laws that have a genuine claim on citizens' obedience would be that the laws must be backed by the general will of the people. The general will, contrasted with the private wills of each citizen, wills only general laws, for the good of each and all, and is necessarily just. It is supposed to be identified by the voting of informed public–spirited citizens, free from divisive factions, in a modestly sized state designed to promote citizens' readiness to subordinate their private wills to the general will. Only in such a state, Rousseau argued, could citizens be morally free—that is, subject only to laws that they give to themselves.

The problem, as Rousseau realized, was that even if such a state could be established, it is unlikely to be stable. Various stabilizing devices, such as minimizing property inequalities, keeping the state small, and promoting civil religion, were to use social forces to promote public–spirited identification with the general will, but Rousseau, like Plato, foresaw how easily an ideal could decline and fall back into moral chaos.

I think that Immanuel Kant's moral philosophy is, in many (but not all) ways, an expression of common–sense morality, and I find it illuminating to look at Kant's *moral* philosophy as a transforming of Rousseau's *political* philosophy without raising the problem of political instability and utopian thinking. The quick version of the story runs as follows. First, unlike Rousseau, Kant was convinced that it is impossible to create an ideal state by depending on the virtue

and good–will of citizens.[7] The spirit of the French Revolution was magnificent, but a legitimate and "rightful" juridical order must have authority to make and enforce laws without relying on the good will of citizens. Principles of justice can only govern the "external" conduct of citizens through universal coercive laws, designed to define and protect property and liberty rights, not to promote the virtue or happiness of citizens. Kant argued that citizens must obey whatever *de facto* government exists, so long as it meets minimal standards of law; and the fact that one's state fails to meet the higher standard of a just Republic does not invalidate its laws or justify disobedience. There is no right to revolution. Political stability must be based on legitimate force, not dependent on the virtue of citizens. Thus, Kant argued that Rousseau's unstable political utopia is not a model for law and justice.

Instead, Kant incorporated ideas central to Rousseau's political philosophy into his moral philosophy. Although (unlike Rousseau) Kant did not think that moral virtue and vice are determined by social forces, he regarded natural human inclinations as neither good nor bad morally.[8] Human nature, he thought, contains the resources for increasing approximation to virtue and human perfection, but at least minimally just laws are necessary for their development. We cannot infer from observing human beings ("examples") what their moral capacities are, but we must attribute a basic moral *capacity* (*Wille*) to all humanity despite the actual corruption we see around us. In place of Rousseau's "general will" as the standard for authoritative public laws, Kant proposed the "good will" of rational autonomous persons as the source of all moral laws. The Kantian good will, like Rousseau's "general will," subordinates merely personal interests to laws legislated from a standpoint of respect and concern for each person. The ideal result is a harmonious community of mutually respecting persons under common *moral* laws, "a kingdom of ends." But the moral laws that good (rational, autonomous) wills could or would legislate remain the basic ethical standard for our conduct independently of whether or not other people obey.[9] The moral laws do not need actual enactment or public endorsement, and wide–spread selfishness and disregard for the standard do not make it less binding or make "moral freedom" impossible. The stability of a public system of laws is no longer

[7] The primary source for Kant's political philosophy is his *Rechtelehre*, the first part of *The Metaphysics of Morals* (Kant 1785).

[8] This, at least, is the mature view expressed in Kant 1793.

[9] Note that this comment refers to the "basic" ethical standard—for Kant the fundamental principles of a metaphysics of morals, but this is compatible with the idea that more specific, context–relative principles incorporate qualifications needed to respond to various kinds of problems posed by the fact that actually people often disregard, disobey, and disagree about ethical norms.

the issue because, as it were, the general will has become an "inner" ideal for individual *ethical* decision–making and *political* stability was to be maintained by forceful "law and order."[10]

Although political stability is not the central issue for Kant, we may suspect that utopian thinking is still a problem because Kantian *moral* theory seems to derive partly from Rousseau's political theory. And this should be a general concern insofar as Kant's moral theory is a re–expression of familiar principles of everyday ethical thinking. The general utopian problem is that it may not be rational or morally obligatory for us to follow ethical principles that would be ideal if everyone were to do so. The cases of lying and torture highlight the problem.

§6. Similar utopian thinking in ethics—in everyday life and in (Kantian) ethics?

(A) Making principles universal:

(1) Ordinary morality: We have all heard our parents or peers attempt to dissuade us from doing something we want to do by saying, "What if everyone did that?" You are supposed to see that it would be bad if everyone did what you propose to do, and so you should not do it. For example, it is argued, "Think what it would be like if everyone refused to pay taxes (or give to charity, vote for an imperfect candidate, etc.). You couldn't choose for that to happen, and so you should pay your taxes (give to charity, vote for the best imperfect candidate, etc.)."

This line of argument asks us to think hypothetically of a possible world unlike ours in that everyone always follows your example—so, if you refuse to pay taxes, they all refuse; if you refuse to give to charity, they all do, etc. If you act on a bad principle, then the result would be a "**dystopia**"—an undesirable imaginary world that extends and exaggerates certain tendencies in our real world to help us to see to them as bad tendencies. The argument indirectly invites us to think of a **utopia** in which everyone adopts and *follows the same good principles* (paying reasonable taxes, giving some to charity, voting, etc.) and so all is well and harmonious. In effect, the argument says, "Don't act on your personal principle that would result in a bad world if everyone followed you. Instead, look for and follow the principles that everyone can share and accept together (consistently,

[10] Rousseau had problems trying to find public observable criteria for identifying the general will (e.g. majority vote under favorable conditions), but Kant states up–front that a good will is a non–empirical ideal of which we may never find an actual example.

with mutually acceptable results)."

Not everyone buys this line of argument, of course. Smart teenagers, for example, have learned to answer back to their parents by saying, "But everyone is not going to do what I'm doing—so why shouldn't I?" And you may think, as I do, that there is something wrong with this line of argument in the following cases. (i) A radical libertarian says, "If everyone refuses to pay taxes, government will shrink to almost nothing—as it should!—and so everyone should refuse to pay taxes." And (ii) extreme pacifists may argue, "If everyone refused to fight, there would be no more wars (etc.)—and so you should always refuse to fight." If the outcome seems unacceptable, we need to question (and maybe re–formulate more carefully) the argument.

(2) Kant: The philosopher Immanuel Kant famously re–expressed the line of argument we have been considering in his first formulation of what he called the Categorical Imperative. This universal law formula (FUL) says that *we should act only on maxims that we can will as universal laws.*[11] Maxims are personal principles that we act on. We are supposed to test whether these are morally worthy maxims by asking ourselves whether we can consistently conceive and choose these personal maxims as principles for everyone to adopt and follow, or at least for everyone to be permitted to adopt and follow. I will spare you the long history of interpretation, critique, and defense of this formula as a moral guide. My main focus, instead, will be on Kant's later formulations of the Categorical Imperative, but I want at least to mention how common uses of Kant's universal law formula seem to invite the charge of utopian thinking—the same sort of objection smart kids raise to the everyday arguments that start with "What if everyone did that?"

What we draw from the formula depends on how we describe the relevant maxim (personal principle) as well as on what it means to will it as universal law and the criteria of what we can and cannot so will. But here is the apparent problem. The test asks us to determine what is permissible here and now from thoughts about the remote possible worlds in which everyone does, can, or may follow us in adopting and acting on our maxim. Unless appropriate qualifications deflect the problem, this seems a paradigm of utopian thinking of the dubious sort.

[11] For simplicity, I shall not distinguish the formula of universal law (FUL) and its variation the formula of universal law of nature (FULN) or discuss the relations between them. See *Groundwork for the Metaphysics of Morals* (Kant 1785: Ak. 4:421, 222) and *Critique of Practical Reason* (Kant 1788: Ak. 5: 31–3, 67–71). This would be important in a study of Kant's texts, but here I am concerned with practical moral guiding function of various interpretations or reconstructions of Kant's ideas.

Consider two sorts of example. *First*, suppose I am considering what to give to relieve people suffering from a particular disaster. If everyone who can do so easily gives a certain modest amount, then the problem will be solved. Call this the "fair share" of each potential giver. (The amount could be calculated progressively, with shares increasingly larger for the wealthier.) So, I adopt the maxim to give exactly my fair share, and if everyone did likewise, the emergency need would be met. In fact, however, many others will give less than their fair share and the few who give more will not compensate for the difference. Acting as would be justifiable if everyone adopted my maxim, arguably, would be wrong in the real world where I know more is needed to meet dire needs.

Suppose now that I adopt maxims never to lie to, steal from, mock, torture, or kill other people. I try to imagine these maxims as "laws" that everyone adopts and acts on, and I may see the resulting possible (fantasy) worlds as better worlds than our real world. And so, I might think that I could (reasonably) choose to bring about these imaginary worlds if I had the power.[12] So, complications aside, it seems that by FUL I should regard myself as at least permitted to act on these unqualified maxims—never to lie to, steal from , mock, torture, or kill other persons.[13] And yet in the real imperfect world, it seems there are many situations, due to the wrongdoing of others, in which, regrettably, lying, stealing, mocking, or killing would be justifiable, for example, because it is the only way to avoid a vastly worse evil or disaster.[14] Since I live in the real world, then, it seems that, contrary to Kant, I should not adopt and act on my absolutely unqualified maxims even if I can will these as maxims for a world where everyone follows suit.

The common Kantian response is to call for a re–description of the maxim, then test the alternative maxims, repeating until the result of applying the

[12] This is because (I am assuming for now) that the only reasons exceptions are considered justifiable is that the non–compliance of others creates problems that require responses that would otherwise be morally objectionable. But see the note below.

[13] Complications might include that extraordinary circumstances could arise because natural forces, not bad human choices, make lying, stealing, etc., justified. We may need to treat even innocent people in normally forbidden ways when natural disasters give us only tragic options. Lying to an innocent and conscientious gate–keeper who, on the basis of reasonable beliefs, refuses to permit us to respond as needed to an immediate catastrophe might be an example. I am not recommending a hybrid Kantian–utilitarian theory, but would suggest that sometimes the most fundamental Kantian grounds for our specific moral prohibitions may also justify exceptions.

[14] One does not have to be a consequentialist to acknowledge the problem here. Even if the justification in the particular case depends on the need to avert bad (e.g. disastrous) consequences, it does not necessarily appeal to principles that make only consequences ultimately relevant. Any sensible non–consequentialist (or "deontological") theory, I think, must admit that at some level and with due qualifications consequences matter. Kant is no exception, despite a bad press.

universal law formula (FUL) matches the moral conclusion that, intuitively, we think correct. But to do this is to give up the idea that this formula of the Categorical Imperative alone is an adequate action guide, whatever its other merits may be.

Another response is to try to make the maxims we test specific enough to reflect relevant differences in context. For example, I could try out the maxim of giving a fair share calculated by dividing the amount needed among those actually willing to contribute, with the wealthier perhaps expected to give more. Or, in the second case, I could test maxims never to lie, steal, mock, or kill *except* when certain exceptional circumstances arise. For example, my maxim might be "I'll never tell a lie unless necessary to save a life or something similarly important." A commonly recognized problem with this proposal is that there seems no adequate way to determine exactly what details should be included in the maxim without presupposing independent moral standards.[15] (A similar problem may apply to the opposite proposal that we use the FUL to test only to the most general, fundamental maxims.[16])

The suspicion of utopian thinking, however, raises a more general and less discussed question. Why, we may wonder, should the working policies we do, can, or would choose for an idealized possible world that is very different from ours be the standard for what we adopt and act on here and now?

(B) Treating each person with dignity that must never be violated:

(1) Ordinary morality: Without thinking of any particular moral *theory*, most people, I think, accept the idea that there are definite and inflexible moral limits on the means that are acceptable for them to use to pursue their ends. Torture, rape, and murder are commonly regarded "off limits" in all cases, even if physical restraint and telling lies are sometimes justified as necessary means to prevent a greater evil.

(2) Kant: Similarly, the idea in Kant's *Groundwork* that resonates with non–specialist readers more than any other is that *we must never treat a person merely as a means, but treat all persons as ends in themselves.* Human beings, Kant implies, have a *dignity* beyond price, regardless of their class, wealth, gender, race,

[15] In *The Practice of Moral Judgment* Herman (1993) tries to address this problem by using FUL to identify relevant considerations that constitute a "deliberative field" for judgment.

[16] See O'Neill, especially "Consistency in Action" (1989: 91–2).

accomplishments, and even moral record.[17] Torture, rape, and murder have seemed to be paradigm examples of failing to respect the basic human dignity of another person.

The problem arises when philosophers construe Kant's humanity formula, or their own Kantian moral theories, as absolutely prohibiting doing the various things that intuitively might be disrespectful of some person.[18] Focusing on ordinary small scale interactions alone, we may regard it as disrespectful to mock people, to use force on them, to lie to them, to manipulate them, disregard their opinions and preferences, to fail to give them life–saving aid, and to impose avoidable risks of harm and death on them. Both intuitively and in Kant's view, to treat people in these ways is wrong under most normal conditions.[19] In a more perfect world, we can imagine, we could and should adhere to these ideals completely, but there are two related problems with interpreting the ideals as giving us *absolute* prohibitions. The *first* is that in the real world the prohibitions, or the grounds for them, *can conflict* in particular cases.[20] The *second* is that the absolute prohibitions *become seriously counter–intuitive* in extraordinary tragic circumstances.

The *first* problem, abstractly stated, is that absolute prohibitions on mockery, use of force, lying, imposing risks, etc., or the reasons for them, can pull in

[17] Kant evidently did not recognize the full implications regarding gender and race, but arguably his idea of human dignity, combined with a more enlightened view of the relevant empirical facts, would be incompatible with repugnant racism and sexism. Bernard Boxill and I have reviewed some of these issues in "Kant and Race" in Boxill (2001: 448–471).

[18] Alan Donagan seems to rely on intuitive understanding of respect in particular interpersonal exchanges even though, unlike most commentators, he aims to present a consistent and coherent system of moral principles analogous to Kant's metaphysics of morals and he does introduce some qualifications. See Donagan 1977 and "Donagan's Kant" in Hill 2000.

[19] This depends, of course, on how broadly we construe "normal." If crime is normal, then the use of force by police would not be regarded as wrong in all normal conditions. Similarly, if war, dealing with the mentally ill, and planning surprise parties are "normal," then intuitive opposition to imposing risks, not treating people as rational, and lying cannot be expected in all normal circumstances. What is considered normal is relative to context, and it serves little purpose here to try to draw clear lines. My main point is just that, although often and presumptively considered wrong in daily life and perhaps absent in a perfect world, the sorts of "disrespectful" treatment listed here are not reasonably construed as always wrong.

[20] Kant grants that "the grounds of duties" may conflict while denying that there can be genuine "conflicts of duty." In principle, then, Kant could admit the qualifications that I suggest below are needed in specific moral precepts, but his readers can hardly fail to notice that he often presents particular precepts as inflexible and unqualified. The initial problem with unresolved conflicts in the grounds of duty is that we might hope that a theory could offer some guidance for choices in those situations. More specific problems arise when a theory resolves the conflicts this way or that.

different directions in some cases.[21] For example, giving life–saving aid may require the use of force or telling a lie, as in Kant's famous story (in "On a supposed right to tell a lie from philanthropy," 1797: *Practical Philosophy,* Ak. 8: 423–30, 605–16). Manipulating their guards may be the only way to rescue people held in slavery.[22] Disregarding an addict's opinions may be the only way to respect his deepest rational preferences. Those in power sometimes have no option but to impose grave risks on one group or another. Although we may doubt whether certain particular cases pose genuine dilemmas, the list of potential conflicts seems endless.

I would argue, then, that everyone, whether a follower of Kant or not, should face up to the potential conflicts between the moral principles that they would like to regard as absolute. We need to think seriously *why* we assume, if we do, that there are no justified exceptions to the principles that we should not torture, rape, or murder and yet there are some justified exceptions to the prohibitions on telling lies and killing, e.g. lying to a Nazi at the door to save his victims from murder and killing in self–defense and just wars. I am not suggesting here that we *should* allow exceptions to the rules against torture, rape, and murder. Rather the point is that to be conscientious we need to reflect on how our principles can conflict, how they work together (like the laws in a good legal system), and what

[21] Note that the intuitive precepts that seem to be implicit in the humanity formula and oppose intuitively disrespectful treatment in ordinary cases, as I presented them, are prohibitions of treating people in certain specified ways, not just prescriptions of ends or values that we should try to promote. For example, "Don't lie" rather than "Lies are bad things (i.e. good to try to prevent)" and "Don't manipulate people" rather than "Try to bring it about that people are not manipulated." From Kant's point of view, and perhaps ours, it is a common fact and not a deep theoretical problem that the good ends we should promote often conflict, requiring us to use good judgment and make reasonable choices about how to balance them. The apparent conflicts of duty I have in mind here are not the often discussed cases where telling a lie, manipulating, or killing someone is expected to prevent more cases of the same from being done by others, but rather cases where one person, it seems, will be doing something prohibited no matter which option she chooses. It is perhaps noteworthy that the examples that come most readily to mind involve conflicts between a prohibition of acts such as lying, stealing, and imposing risks of harm and failures to give desperately needed aid. Perhaps anticipating this, Kant classified the duty to give aid under the imperfect duty of beneficence, giving it a lower priority in case of conflict with "perfect" duties to avoid lies, etc. This way he could maintain both strict prohibitions on lying, etc., and a significant duty to give aid without allowing a conflict of strict duties in his system of moral principles. This way of handling the problem, of course, raises the second utopian suspicion I mentioned—namely, that adherence to the ideal code for perfect rule–followers does not allow for exceptions intuitively justified by extraordinary circumstances in the real world.

[22] Note that this could be so even if the guards were conscientious persons of good will if they could only be moved by otherwise impermissible manipulation because they thought they were holding only murderers and there was no time or means to convince them otherwise.

grounds we have for making exceptions in some cases and not in others without being arbitrary (like flipping a coin or choosing whatever you feel like).

Similarly, moral philosophers in the Kantian tradition have a responsibility to try to work systematically to reconcile conflicts among the precepts that seem to follow from Kant's humanity formula—the idea that every human being has a dignity that must never be violated.[23] A default position is just to acknowledge that the precepts are defeasible, and leave further resolution to judgment in particular cases. We should perhaps also admit that the resources of our moral theory, and perhaps any theory, are incapable of answering all moral questions.[24] We can aspire, however, to make some progress in (systematic normative) ethics by trying to work out more specifically how certain principles should be qualified and how, if at all, they can be ranked relative to each other. Kant's perspective of ideal moral legislators in a kingdom of ends, I suggest later, can be seen as a deliberative framework for this sort of constructive review of principles, priorities, and possible exceptions.[25]

The *second* problem, for both ordinary morality and Kantian defenders of human dignity, is that if you take your principles as absolute and without exception, then although they might be appropriate for a more perfect world (a utopia), the implications of construing them so strictly in real extraordinary tragic circumstances can seem intolerable. Most people, I find, disagree with Kant's inflexible prohibition on lying, and his own story of lying to a murderer to save a friend serves for them as a counter–example to his position. In a perfect world, perhaps, no one would ever tell serious lies, but in our imperfect world it seems fully justified to save a friend from murder by means of a deliberate lie. Mockery, or expressing utter contempt for a person, seems a paradigm of disrespect, but even here an absolute prohibition is unwarranted. Some might think that the world's most evil villains deserve such expressions of contempt,[26] but, that aside, consider a case of mocking an innocent person. Imagine a

[23] As mentioned above (note 21) Kant does try to reduce conflicts in his system of principles by classifying some as merely "imperfect" and others as "perfect," but to eliminate such conflicts was more an ambition than an achievement and he lays out his priorities among principles with relatively little supporting argument.

[24] "Moral Dilemmas, Gaps, and Residues" in Hill 2002 argues for accepting that there are "gaps" in Kantian moral theory and that this is not altogether a bad thing.

[25] Kant 1797–8 proposed some partial ranking in by distinguishing various types of duty (perfect v. imperfect, duties of justice v. duties of virtue, etc.). Similarly, much of Rawls 1971 is an attempt to argue for some ranking among considerations of justice rather than simply labeling them all as "prima facie" or "pro tanto" principles to be applied intuitively case by case.

[26] This seems contrary, however, to Kant's position in *The Metaphysics of Morals* (1797–8: Ak. 6: 463, 209).

situation of an undercover agent who needs to protect her "cover" in order to rescue a victim being mocked, say with racial epithets, by a vicious gang planning to do even worse to this victim and others.[27] If the agent refuses to join in the mockery now, she will be discovered and lose all chance to rescue the victim and others. Sticking inflexibly to the precept "Don't mock people" is short–sighted. A more reasonable understanding of what respect for persons requires would admit exceptions. The set of principles that best respects all persons in a complex and imperfect world, I suggest, must allow qualifications to many specific precepts that are normally appropriate and would be in a more ideal world.

(C) Respecting principles (and exceptions) that reasonable person would accept:

(1) Ordinary morality: Everyday thinking about ethics is not all about absolutely inflexible rules. In every profession, and not just on TV, people face emergencies and crises where they realize the best thing they can do is "to bend the rules" and "make an exception." We say, "Everyone would agree that *normally* we shouldn't do this sort of thing, but any reasonable person should understand that in the circumstances it was justified." Here implicitly the hypothetical agreement of reasonable people sets the standard for both the norm and the justified exception, and "reasonable people" are assumed to understand the problem, to take into account its complex context, including all of the values and people involved. We typically assume that all reasonable people agree that some norms allow exceptions but others do not—for example, we may break a minor promise to save a life, but vindictive torture, rape, and murder is never justified. The problem is, how can we draw the line and justify our decision to others?

(2) Kant: In the *Groundwork* Kant suggests that ideas from his previous formulations of the Categorical Imperative are combined in a fruitful way in the principle that *we should always conform to the laws of a possible moral commonwealth ("a kingdom of ends")* (1785: Ak. 4:433–40, 233–40).[28] This is conceived as moral

[27] Kant opposed telling lies (1797: Ak. 8: 423–30, *Practical Philosophy*, 605–16) and the use of spies (1795: 71), so he would no doubt have disapproved of the undercover project in the first place, despite its potentially desirable consequences. The "Kantian" position I recommend deviates from Kant's own views at this point.

[28] Over the years I have often, with repetition to a now embarrassing extent, proposed an interpretation or reconstruction of Kant's "kingdom of ends" as an apparent improvement on the formula of universal law and the humanity formula and as a promising idea for normative moral theory. The Kantian idea naturally invites a charge of utopian thinking and so may be worth discussing here. The main idea,

union of rational and autonomous persons under common laws of which they are both *authors* and *subjects*.[29] In legislating the members are governed by practical reason and so, presumably, respect each other as "ends in themselves." They have "autonomy of the will" and so they "abstract from personal differences" and legislate only laws only that all can endorse. As a formulation of the Categorical Imperative, the principle expresses the general moral requirement to conform to the more specific (or mid–level) principles that would be "legislated" or endorsed by all the members of this ideal moral union. As a summary expression of a Kantian moral theory, the general idea is that we should view substantive moral principles as authoritative because they would be endorsed by and for all rational and reasonable people who, while setting aside their individual differences, aim find principles that harmonize their personal ends and respect the dignity of each.[30] From these abstract ideas, I have suggested elsewhere, we might try to develop a Kant*ian* deliberative framework beyond Kant's texts, as a perspective for assessing specific moral principles, trying to reconcile potential conflicts, and deliberating about possible exceptions.[31]

The idea of moral legislation here has some apparent advantages over FUL or FHE taken by themselves.[32] For example, systematic thinking is encouraged about how our moral norms work together. Without conflating law and ethics, we can see by analogy that we need to evaluate sets of rules and principles together. What might make sense as a "universal law" in the context of certain laws might be unacceptable in the context of other laws. Thinking systematically also exposes potential conflicts and invites serious wide–ranging reflection on how they should be managed. The analogy with law also suggests that legislators can continue to affirm constitutional essentials while allowing reasonable limits and exceptions to the specific rules they adopt. This enables us to address the

admittedly an extension and modification of Kant's explicit views, is described briefly in Hill 1992, chapter 11, and developed further in Hill 2000, chapters. 2, 4, and 8. Some objections to the use of the idea of hypothetical agreement in the legislative model are discussed in Hill 2002, chapter 3. There is further discussion in Hill 2003, 2006 and 2008.

29 The exception is that the "head," lacking human needs and imperfections, is not "subject" to the laws but necessarily conforms. This is a reference to God, or a "holy will," who formally shares a traditional characteristic of sovereignty (author, not subject to the laws) but necessarily wills the same laws as the other rational autonomous legislating members and so is not said to be "bound" or under "obligation" because of them.

30 This summary conception might help, as Kant says, by using an analogy, bring the moral law "nearer to feeling" and "closer to intuition" (Kant 1785: Ak. 4: 437, 237).

31 See also Rawls 1971, chapter on Duties and Obligations of Individuals.

32 These putative advantages are noted in "The Kingdom of Ends" in Hill (1992: 58–66).

problem that the absolute prohibitions often drawn from FHE are counter–intuitive in some extreme cases. In light of the problems generated by attempts to test *maxims* by FUL, the legislative perspective is promising, too, because its application does not require us to pick out one most salient act–description to determine what is permissible. If, under any true description, acts are incompatible with the system of "legislated" moral principles, they would be forbidden, otherwise not.

Even though the kingdom of ends formula addresses utopian objections to Kant's other formulas of the Categorical Imperative, it poses its own dangers of utopian thinking. Let us consider, all too briefly, three of these.

(a) ***Not everyone will actually follow the best the rules.*** The most obvious objection is that ideal rules for perfect rule–followers may be foolish and dangerous when used to guide us in our world of imperfect rule–followers. The problem is akin to the concern that Rawls' principles designed for a "well–ordered" society are unfit for societies marked by deep conflicts. If the members of the kingdom of ends are conceived as always conscientiously following the laws that they make, then their laws may not be applicable to us.

To meet the objection, we may interpret or modify the model as a deliberative perspective designed to find the moral laws that *ideal legislators* would adopt for a world of *non–ideal moral agents*. In other words, we are to ask what, in our best possible deliberations, we would endorse as principles for ourselves and others under all sorts of imperfect human conditions. The principles must *apply* to human beings as we are, which is often ignorant, foolish, biased, weak, and corrupt; but to be morally authoritative *law–making* must be appropriately informed, wise, unbiased, consistent, and well–intended. And so we need to consider what principles we would adopt if we were ideal deliberators but taking into account human frailty and corruption.[33]

[33] This response may seem to "compromise" and "water–down" our moral principles in order to accommodate our failings in a way that Kant would resist, but arguably the response fits with Kant's understanding of morality. Moral imperatives and obligation only apply to finite persons who have "imperfect wills," subject to the limitations and vulnerabilities of human sensibility. Pure practical reason in each person is supposed to be "legislating" for us, but what it legislates is a system of principles (a "metaphysics of morals") appropriate for human capacities and limitations. Those who make exception to such principles in order to make excuses for their avoidable failings are illegitimately compromising the principles, but in theory the principles are reasonably applicable for human conditions that can vary widely in their particular kinds and degrees of imperfection.

(b) ***Even conscientious people disagree.***[34] An ideal model might be regarded as utopian is if it turned out to have few or no implications for how we should act, and one might suspect that this is true of the Kantian legislative model because even the best moral deliberators often disagree in their judgments. If the theory says that the correct or fully justified moral principles are just those that *all* ideal moral deliberators would adopt under the specified conditions, then we cannot know what such principles (if any) are unless we can know that ideal deliberators would come to agreement on them. Where there is potential disagreement as far as we can tell, it seems the model can give no guidance. And the values and priorities that Kantians can plausibly derive from the idea of human dignity are surely insufficiently precise and determinate to guarantee agreement among those who use them to assess systems of principles. Is this a decisive objection?

A possible response is, first, that the stipulated values are not *so* indeterminate that disagreement can be expected on all important matters. After all, the requirement is not that every actual person, regardless of their values and deliberative point of view, will actually agree. (One can challenge those values and the Kantian deliberative perspective, but that is another issue.) But, this aside, let us grant that it is plausible to suppose that often even the best deliberators will disagree especially on relatively specific principles and the exceptions that they may admit. One suggestion that acknowledges potential reasonable disagreements but preserves the action–guiding function of the Kantian model is this: treat the Kantian deliberative perspective as a guide for conscientious decision–making and theoretical aspiration, with complete convergence of judgment among ideal deliberators as a criterion of full justification or "moral truth" rather than a necessary condition for regarding principles as valid for action–guiding purposes. That is, the main point is that individuals should follow those principles that, in their best possible judgment, they would commend to anyone as to–be–endorsed from the Kantian deliberative perspective, whether or not they are fully confident that everyone taking that perspective would agree. This does not ignore the opinions of others, but on the contrary puts pressure on us to challenge our initial judgments by consulting with them—otherwise our judgments are not even conscientious.

(c) Making exceptions is a "slippery slope." If you begin to make exceptions to common moral principles, we may find it easier and easier to do so until you find at last that you really have no principles. (A river dam that is full of holes soon ceases to be a dam.) Admittedly this is a practical problem, but sometimes

[34] My discussion of disagreement and conflict draws from Hill 2016.

philosophers raise a deeper, theoretical problem that I call "the slippery slope problem" for anyone who thinks that there is more to consider in ethics than just the amount of happiness or welfare that your act brings about. Suppose you say it is wrong to torture. Then a natural challenge is, "But what if a little bit of torture to known malicious terrorist will save a million innocent lives?" If you say, "All right, in that case," the next question is "How about if it will save three quarters of a million innocent lives?" If you say, "Well, yes, this case too would be an exception," then your challenger will ask, "How about a quarter of a million?", then "100, 000?", then "1, 000?", then "100?", then "10? You can see where this is going. The critic's suggestion is that if your seemingly firm stand on principle gives way *merely because it is expected to result in more happiness (or aggregate welfare)*, then it seems that, after all, this is your ultimate standard—just doing whatever has the best consequences in terms of promoting "the greatest happiness."

Neither ordinary morality nor Kantian ethics needs to accept this challenge as a valid criticism, I think. There is a difference between being bald and not being bald even though, as we lose one hair at a time there is no precise point at which we could say definitely "NOW you are bald." More important, my suggestion here has been that both ordinary morality and Kantian ethics, properly understood, take into account more than the effect of our acts on the general happiness. The ideas of human dignity and respecting principles that can be justified to others do not serve as quick and easy guides to moral decisions in every case, but they call for good judgment and thoughtful reflection on how we can best respect our deepest values in an imperfect and cruel world, where they often pull us tragically in different directions. The key is to make sure that the exceptions are justifiable by the same ultimate values, such as human dignity and moral community, that stand behind moral principles in normal, non–emergency cases.

§7. Concluding Note

I began with the troublesome cases of (A) telling a lie to save a friend from a murderer and (B) torturing a suspected terrorist to get information to stop a bombing. Are there justified exceptions to lying and torture in Kantian ethics?

In the first case, I think, most of us would agree that telling a lie is justified, and what is troublesome (at least to those sympathetic to Kant's philosophy) is that Kant himself said that telling the lie would be wrong! In sum, my response is this. A reasonable core of Kant's ethics can be separated from his own extreme personal opinions about particular cases. And this reasonable core, as I call it, does not condemn lying to the murderer. Why? Well, to start, one can

consistently and sensibly will a maxim (or policy) of not telling lies *except* in certain kinds of circumstances—for example, when the lie is to someone making immoral threats and lying is the only way to prevent grave harm to an innocent person. Also, lying to someone to deflect him from a murderous project is hardly a violation of his human dignity. Arguably free, equal, and reasonable moral "legislators" would adopt a *general presumptive* rule against telling lies in ordinary conditions, for example, to take advantage of others, but they could recognize that the rules about lying and truth–telling that best affirm the dignity of each person would admit of exceptions when we take into account imperfect real world conditions. Arguably the exceptions here are justified by the very principle that normally demands truth–telling.

The case of torture in interrogation is different, I think. The case is troublesome to most of us, I expect, because we see two important moral principles apparently in conflict. On the one hand, though our primary responsibility is to make sure that we ourselves do not wrongfully assault, murder, coerce, cheat, and debase other people, we have responsibility to try (by moral means) to prevent others from committing harmful and immoral acts. *So, we have a reason to try to prevent terrorist suspects from carrying out their bombing plot.* On the other hand, interrogation by torture deliberately inflicts intense pain and humiliation on a captive who may be conscientious or even innocent, to break his will and turn him against himself in an attempt to extract information that he may not have. We use the tortured captive as a means to get information but in so doing, it seems, we assault his dignity as a human being. *So we also have strong reason to avoid the use of torture.* How would reasonable Kantian moral legislators think about the conflict here? All too quickly, I would say that a review of the history of the use and abuse of torture would give them reason to condemn torture in the strongest terms and *for all reasonably foreseeable circumstances.* Making exception for some extraordinary case would not be *inconceivable*, but in the real world the problem is not what is suggested by fanciful stories about the "ticking bomb." A major difference between these fictional stories and real life is that it is stipulated in the stories *we know* that *in fact* the torture is absolutely necessary and will be completely successful in preventing a bombing. In the real world we do not have such knowledge, and moral principles are made for fallible people who must act on estimates and guesses. Importantly, to be effective the relevant principles must be publicly promulgated, learned, and enforced. From the perspective of Kantian legislators introducing an exception to the prohibition of torture, I think, would be hard to justify. Would they, as legislators committed to the dignity of humanity in each person, adopt the universal but *qualified* principle, "Don't interrogate with torture unless you *think* your using torture will

prevent others from causing serious harm to innocent people"? The dangers of an exception like this should be evident, especially given there is much data now to confirm what professional interrogators have often said, namely, torture is *not* the most effective way to get information from captives. The main Kantian reasons against the practice of torture, of course, stem from the horrible and degrading way it treats a human being—the empirical evidence regarding its ineffectiveness merely helps to set aside facile attempts to justify exceptions. Anyway, this is the way one could try to respond to the problems, but I should emphasize that my reasonable core of Kantian ethics acknowledges that reasonable people can disagree about how fundamental principles should be applied to hard cases.

REFERENCES

BOXILL, Bernard (ed.) (2001). *Race and Racism.* Oxford: Oxford University Press.

DONAGAN, Alan (1977). *The Theory of Morality.* Chicago: University of Chicago Press.

ENGELS, Friedrich (1878). *Landmarks of Scientific Socialism: Anti–Duhring.* Translated by Austin Lewis. Chicago: Charles H. Kerr & Company, 1907.

HERMAN, Barbara (1993). *The Practice of Moral Judgment.* Cambridge, Massachusetts and London: Harvard University Press.

HILL, Thomas (1992). *Dignity and Practical Reason in Kant's Moral Theory.* Ithaca: Cornell University Press.

HILL, Thomas (2000). *Respect, Pluralism, and Justice: Kantian Perspectives.* Oxford: Oxford University Press.

HILL, Thomas (2002). *Human Welfare and Moral Worth: Kantian Perspectives.* Oxford: Oxford University Press.

HILL, Thomas (2003). Treating Criminals as Ends in Themselves. *The Annual Review of Law and Ethics* Band 11: 17–36.

HILL, Thomas (2006). *Lindlay Lecture: The Importance of Moral Rules and Principles.* Lawrence: University of Kansas Press.

HILL, Thomas (2008). Moral Construction as a Task: Sources and Limits. *Social Philosophy and Policy* 25 (1): 214–236.

HILL, Thomas (2016). Human Dignity and Tragic Choices. *Proceedings and Addresses of the American Philosophical Association* 89: 74–94.

KANT, Immanuel (1785). *Groundwork for the Metaphysics of Morals.* Translated by Arnuf Zweig and edited by Thomas E. Hill, Jr. and Arnulf Zweig. Oxford: Oxford University Press, 2002.

KANT, Immanuel (1788). *Critique of Practical Reason.* Translated by Mary Gregor. Cambridge: Cambridge University Press, 1997.

KANT, Immanuel (1793–4). *Religion within the Boundaries of Mere Reason.* Translated and edited by Allen Wood and G. Di Giovanni. Cambridge: Cambridge University Press, 1998.

KANT, Immanuel (1795). *Toward Perpetual Peace.* In *Toward Perpetual Peace and Other Writings on Politics, Peace, and History,* translated by David L. Colclasure and edited by Pauline Kleingeld. New Haven: Yale University Press, 2006.

KANT, Immanuel (1797). *On a supposed right to tell a lie from philanthropy.* In *Practical Philosophy,* translated and edited by Mary Gregor. Cambridge: Cambridge

University Press, 1996, pp. 605–6.

KANT, Immanuel (1797–8). *The Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1996.

MARX, Karl (1844). Introduction to Critique of Hegel's Philosophy of Right. *Deutsch–Französische Jahrbucher*. Paris.

MORE, Thomas (1516). *Utopia*. Translated and edited by Paul Turner. London: Penguin Books, 2003.

O'NEILL, Onora (1989). *Constructions of Reason: Exploring Kant's Practical Philosophy*. Cambridge: Cambridge University Press.

PLATO. *The Republic*. Translated by G.M.A. Grube and revised by C.D.C. Reeve. Indianapolis: Hackett Publishing Company, 1992.

RAWLS, John (1971). *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.

RAWLS, John (1996). *Political Liberalism*. New York: Columbia University Press.

RAWLS, John (2001). *The Law of Peoples*. Cambridge, Massachusetts: Harvard University Press.

ROUSSEAU, Jean–Jacques (1755). *Discourse on the Origin of Inequality*. Translated by Franklin Philip. Oxford: Oxford University Press, 1994.

ROUSSEAU, Jean–Jacques (1762). *The Social Contract*. Translated by Maurice Cranston. London: Penguin Books, 1968.

Kantian Ethics and Utopian Thinking

Is Kantian Ethics guilty of utopian thinking? First, potentially good and bad uses of utopian ideals are distinguished, then an apparent path is traced from Rousseau's unworkable political ideal to Kant's ethical ideal. Three versions of Kant's Categorical Imperative (and their counter–parts in common moral discourse) are examined briefly for the ways that they may raise the suspicion that they manifest or encourage bad utopian thinking. In each case Kantians have available responses to counter the suspicion, but special attention is directed to the version that says "Act on the maxims of a universally law–giving member of a kingdom of ends." Interpretations vary, but here primary focus is on one contemporary reconstruction and development of this central idea. Several objec–tions suggest that this idea encourages bad utopian thinking are briefly addressed: that we cannot count on everyone to follow ideal rules, that even conscientious people disagree in their moral judgments, and that theories that allow exceptions to familiar moral rules create a "slippery slope" to moral chaos.

Keywords: Kantian Ethics · Categorical Imperative · Exceptions · Utopias · Utopian Thinking.

Ética kantiana y pensamiento utópico

La ética de Kant ¿es culpable de pensamiento utópico? Primero se hace una distinción entre los usos potencialmente buenos y malos de los ideales utópicos. Luego se traza lo que parece ser una ruta del ideal político inviable al ideal ético de Kant. Se examinan brevemente tres versiones del imperativo categórico (y sus contrapartes en el discurso moral común) en cuanto a las maneras en que podrían hacer surgir la sospecha de que podrían manifestar o inspirar un pensar utópico malo. Un kantiano dispone en cada uno de estos casos de respuestas para argumentar contra esta sospecha. Se presta atención particular, sin embargo, a la versión que dice «actúa sobre la máxima de un legislador universal de un reino de fines». Las interpretaciones varían, pero el enfoque principal se pone aquí sobre una reconstrucción y un desarrollo contemporáneos de esta idea central. Se tratan brevemente varias objeciones que sugieren que esta idea inspira un pensar utópico malo: que no podemos confiar de que todo el mundo sigue reglas ideales, que hasta gente escrupulosa discrepa en sus juicios morales y que las teorías que dan lugar a excepciones de reglas morales conocidas generan un caos moral de «situación resbalos».

Palabras Clave: Ética kantiana · Imperativo categórico · Excepciones · Utopías · Pensar utópico.

THOMAS E. HILL, JR. is Kenan Professor, *Emeritus*, at the University of North Carolina at Chapel Hill. His Ph.D. is from Harvard University. His main interests are Kantian ethics and political philosophy. His books are *Autonomy and Self–Respect* (Cambridge University Press, 1991), *Dignity and Practical Reason in Kant's Moral Theory* (Cornell University Press, 1992), *Respect, Pluralism, and Justice* (Oxford University Press, 2000), *Human Welfare and Moral Worth* (Oxford University Press, 2002), *Virtue, Rules, and Justice* (Oxford University Press, 2012). He co–edited (with Arnulf Zweig) *Kant: Groundwork for the Metaphysics of Morals* (Oxford University Press, 2002).

INFORMACIÓN DE CONTACTO | CONTACT INFORMATION: Department of Philosophy, University of North Carolina, Chapel Hill, N.C. 27599, U.S.A. e–mail (✉): thill@email.unc.edu

HISTORIA DEL ARTÍCULO | ARTICLE HISTORY

Received: 3–*October–2019*; Accepted: *21–December–2019*; Published Online: *31–December–2019*

COMO CITAR ESTE ARTÍCULO | HOW TO CITE THIS ARTICLE

Hill, Thomas (2019). «Kantian Ethics and Utopian Thinking». *Disputatio. Philosophical Research Bulletin* 8, no. 11: pp. 505–529.

© Studia Humanitatis – Universidad de Salamanca 2019