

8. Kant's Ethics

It is rare for a philosopher in any era to make a significant impact on any single topic in philosophy. For a philosopher to impact as many different areas as Kant did is extraordinary. His [ethical theory](#) has been as influential as, if not more influential than, his work in epistemology and metaphysics. Most of Kant's work on ethics is presented in two works. *The Foundations of the Metaphysics of Morals* (1785) is Kant's "search for and establishment of the supreme principle of morality." In *The Critique of Practical Reason* (1787) Kant attempts to unify his account of practical reason with his work in the *Critique of Pure Reason*. Kant is the primary proponent in history of what is called deontological ethics. Deontology is the study of duty. On Kant's view, the sole feature that gives an action moral worth is not the outcome that is achieved by the action, but the motive that is behind the action. The categorical imperative is Kant's famous statement of this duty: "Act only according to that maxim by which you can at the same time will that it should become a universal law."

a. Reason and Freedom

For Kant, as we have seen, the drive for total, systematic knowledge in reason can only be fulfilled with assumptions that empirical observation cannot support. The metaphysical facts about the ultimate nature of things in themselves must remain a mystery to us because of the spatiotemporal constraints on sensibility. When we think about the nature of things in themselves or the ultimate ground of the empirical world, Kant has argued that we are still constrained to think through the categories, we cannot think otherwise, but we can have no knowledge because sensation provides our concepts with no content. So, reason is put at odds with itself because it is constrained by the limits of its transcendental structure, but it seeks to have complete knowledge that would take it beyond those limits.

Freedom plays a central role in Kant's ethics because the possibility of moral judgments presupposes it. Freedom is an idea of reason that serves an indispensable practical function. Without the assumption of freedom, reason cannot act. If we think of ourselves as completely causally determined, and not as uncaused causes ourselves, then any attempt to conceive of a rule that prescribes the means by which some end can be achieved is pointless. I cannot both think of myself as entirely subject to causal law and as being able to act according to the conception of a principle that gives guidance to my will. We cannot help but think of our actions as the result of an uncaused cause if we are to act at all and employ reason to accomplish ends and understand the world.

So reason has an unavoidable interest in thinking of itself as free. That is, theoretical reason cannot demonstrate freedom, but practical reason must assume for the purpose of action. Having the ability to make judgments and apply reason puts us outside that system of causally necessitated events. "Reason creates for itself the idea of a spontaneity that can, on its own, start to act—without, i.e., needing to be preceded by another cause by means of which it is determined to action in turn, according to the law of causal connection," Kant says. (A 533/B 561) In its intellectual domain, reason must think of itself as free.

It is dissatisfying that he cannot demonstrate freedom; nevertheless, it comes as no surprise that we must think of ourselves as free. In a sense, Kant is agreeing with the common sense view that how I choose to act makes a difference in how I actually act. Even if it were possible to give a predictive empirical account of why I act as I do, say on the grounds of a functionalist psychological theory, those considerations would mean nothing to me in my deliberations. When I make a decision about what to do, about which car to buy, for instance, the mechanism at work in my nervous system makes no difference *to me*. I still have to peruse *Consumer Reports*, consider my options, reflect on my needs, and decide on the basis of the application of general principles. My first person perspective is unavoidable, hence the deliberative, intellectual process of choice is unavoidable.

b. The Duality of the Human Situation

The question of moral action is not an issue for two classes of beings, according to Kant. The animal consciousness, the purely sensuous being, is entirely subject to causal determination. It is part of the causal chains of the empirical world, but not an originator of causes the way humans are. Hence, rightness or wrongness, as concepts that apply to situations one has control over, do

not apply. We do not morally fault the lion for killing the gazelle, or even for killing its own young. The actions of a purely rational being, by contrast, are in perfect accord with moral principles, Kant says. There is nothing in such a being's nature to make it falter. Its will always conforms with the dictates of reason. Humans are between the two worlds. We are both sensible and intellectual, as was pointed out in the discussion of the first *Critique*. We are neither wholly determined to act by natural impulse, nor are we free of non-rational impulse. Hence we need rules of conduct. We need, and reason is compelled to provide, a principle that declares how we ought to act when it is in our power to choose

Since we find ourselves in the situation of possessing reason, being able to act according to our own conception of rules, there is a special burden on us. Other creatures are *acted upon* by the world. But having the ability to choose the principle to guide our actions makes us *actors*. We must exercise our will and our reason to act. Will is the capacity to act according to the principles provided by reason. Reason assumes freedom and conceives of principles of action in order to function.

Two problems face us however. First, we are not wholly rational beings, so we are liable to succumb to our non-rational impulses. Second, even when we exercise our reason fully, we often cannot know which action is the best. The fact that we can choose between alternate courses of actions (we are not determined to act by instinct or reason) introduces the possibility that there can be better or worse ways of achieving our ends and better or worse ends, depending upon the criteria we adopt. The presence of two different kinds of object in the world adds another dimension, a moral dimension, to our deliberations. Roughly speaking, we can divide the world into beings with reason and will like ourselves and things that lack those faculties. We can think of these classes of things as ends-in-themselves and mere means-to-ends, respectively. Ends-in-themselves are **autonomous** beings with their own agendas; failing to recognize their capacity to determine their own actions would be to thwart their freedom and undermine reason itself. When we reflect on alternative courses of action, means-to-ends, things like buildings, rocks, and trees, deserve no special status in our deliberations about what goals we should have and what means we use to achieve them. The class of ends-in-themselves, reasoning agents like ourselves, however, do have a special status in our considerations about what goals we should have and the means we employ to accomplish them. Moral actions, for Kant, are actions where reason leads, rather than follows, and actions where we must take other beings that act according to their own conception of the law into account.

c. The Good Will

The will, Kant says, is the faculty of acting according to a conception of law. When we act, whether or not we achieve what we intend with our actions is often beyond our control, so the morality of our actions does not depend upon their outcome. What we can control, however, is the will behind the action. That is, we can will to act according to one law rather than another. The morality of an action, therefore, must be assessed in terms of the motivation behind it. If two people, Smith and Jones, perform the same act, from the same conception of the law, but events beyond Smith's control prevent her from achieving her goal, Smith is not less praiseworthy for not succeeding. We must consider them on equal moral ground in terms of the will behind their actions.

The only thing that is good without qualification is the good will, Kant says. All other candidates for an intrinsic good have problems, Kant argues. Courage, health, and wealth can all be used for ill purposes, Kant argues, and therefore cannot be intrinsically good. Happiness is not intrinsically good because even being worthy of happiness, Kant says, requires that one possess a good will. The good will is the only unconditional good despite all encroachments. Misfortune may render someone incapable of achieving her goals, for instance, but the goodness of her will remains.

Goodness cannot arise from acting on impulse or natural inclination, even if impulse coincides with duty. It can only arise from conceiving of one's actions in a certain way. A shopkeeper, Kant says, might do what is in accord with duty and not overcharge a child. Kant argues, "it is not sufficient to do that which should be morally good that it conform to the law; it must be done for the sake of the law." (*Foundations of the Metaphysics of Morals*, Akademie pagination 390) There is a clear moral difference between the shopkeeper that does it for his own advantage to

keep from offending other customers and the shopkeeper who does it from duty and the principle of honesty. (*Ibid.*, 398) Likewise, in another of Kant's carefully studied examples, the kind act of the person who overcomes a natural lack of sympathy for other people out of respect for duty has moral worth, whereas the same kind act of the person who naturally takes pleasure in spreading joy does not. A person's moral worth cannot be dependent upon what nature endowed them with accidentally. The selfishly motivated shopkeeper and the naturally kind person both act on equally subjective and accidental grounds. What matters to morality is that the actor think about their actions in the right manner.

We might be tempted to think that the motivation that makes an action good is having a positive goal—to make people happy, or to provide some benefit. But that is not the right sort of motive, Kant says. No outcome, should we achieve it, can be unconditionally good. Fortune can be misused, what we thought would induce benefit might actually bring harm, and happiness might be undeserved. Hoping to achieve some particular end, no matter how beneficial it may seem, is not purely and unconditionally good. It is not the effect or even the intended effect that bestows moral character on an action. All intended effects “could be brought about through other causes and would not require the will of a rational being, while the highest and unconditional good can be found only in such a will.” (*Ibid.*, 401) It is the possession of a rationally guided will that adds a moral dimension to one's acts. So it is the recognition and appreciation of duty itself that must drive our actions.

d. Duty

What is the duty that is to motivate our actions and to give them moral value? Kant distinguishes two kinds of law produced by reason. Given some end we wish to achieve, reason can provide a *hypothetical imperative*, or rule of action for achieving that end. A hypothetical imperative says that *if* you wish to buy a new car, *then* you must determine what sort of cars are available for purchase. Conceiving of a means to achieve some desired end is by far the most common employment of reason. But Kant has shown that the acceptable conception of the moral law cannot be merely hypothetical. Our actions cannot be moral on the ground of some conditional purpose or goal. Morality requires an unconditional statement of one's duty.

And in fact, reason produces an absolute statement of moral action. The moral imperative is unconditional; that is, its imperative force is not tempered by the conditional “*if* I want to achieve some end, *then* do X.” It simply states, do X. Kant believes that reason dictates a *categorical imperative* for moral action. He gives at least three formulations of the Categorical Imperative.

- . “Act only according to that maxim by which you can at the same time will that it should become a universal law.” (*Ibid.*, 422)
- . “Act as though the maxim of your action were by your will to become a universal law of nature.” (*Ibid.*)
- . Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only.” (*Ibid.*, 429)

What are Kant's arguments for the Categorical Imperative? First, consider an example. Consider the person who needs to borrow money and is considering making a false promise to pay it back. The maxim that could be invoked is, “when I need of money, borrow it, promising to repay it, even though I do not intend to.” But when we apply the universality test to this maxim it becomes clear that if everyone were to act in this fashion, the institution of promising itself would be undermined. The borrower makes a promise, willing that there be no such thing as promises. Thus such an action fails the universality test.

The argument for the first formulation of the categorical imperative can be thought of this way. We have seen that in order to be good, we must remove inclination and the consideration of any particular goal from our motivation to act. The act cannot be good if it arises from subjective impulse. Nor can it be good because it seeks after some particular goal which might not attain the good we seek or could come about through happenstance. We must abstract away from all hoped for effects. If we remove all subjectivity and particularity from motivation we are only left with will to universality. The question “what rule determines what I ought to do in this situation?” becomes “what rule ought to universally guide action?” What we must do in any situation of moral choice is act according to a maxim that we would will everyone to act according to.

The second version of the Categorical Imperative invokes Kant's conception of nature and draws on the first *Critique*. In the earlier discussion of nature, we saw that the mind necessarily

structures nature. And reason, in its seeking of ever higher grounds of explanation, strives to achieve unified knowledge of nature. A guide for us in moral matters is to think of what would not be possible to will universally. Maxims that fail the test of the categorical imperative generate a contradiction. Laws of nature cannot be contradictory. So if a maxim cannot be willed to be a law of nature, it is not moral.

The third version of the categorical imperative ties Kant's whole moral theory together. Insofar as they possess a rational will, people are set off in the natural order of things. They are not merely subject to the forces that act upon them; they are not merely means to ends. They are ends in themselves. All means to an end have a merely conditional worth because they are valuable only for achieving something else. The possessor of a rational will, however, is the only thing with unconditional worth. The possession of rationality puts all beings on the same footing, "every other rational being thinks of his existence by means of the same rational ground which holds also for myself; thus it is at the same time an objective principle from which, as a supreme practical ground, it must be possible to derive all laws of the will." (*Ibid.*, 429)

9. Kant's Criticisms of Utilitarianism

Kant's criticisms of utilitarianism have become famous enough to warrant some separate discussion. Utilitarian moral theories evaluate the moral worth of action on the basis of happiness that is produced by an action. Whatever produces the most happiness in the most people is the moral course of action. Kant has an insightful objection to moral evaluations of this sort. The essence of the objection is that utilitarian theories actually devalue the individuals it is supposed to benefit. If we allow utilitarian calculations to motivate our actions, we are allowing the valuation of one person's welfare and interests in terms of what good they can be used for. It would be possible, for instance, to justify sacrificing one individual for the benefits of others if the utilitarian calculations promise more benefit. Doing so would be the worst example of treating someone utterly as a means and not as an end in themselves.

Another way to consider his objection is to note that utilitarian theories are driven by the merely contingent inclination in humans for pleasure and happiness, not by the universal moral law dictated by reason. To act in pursuit of happiness is arbitrary and subjective, and is no more moral than acting on the basis of greed, or selfishness. All three emanate from subjective, non-rational grounds. The danger of utilitarianism lies in its embracing of baser instincts, while rejecting the indispensable role of reason and freedom in our actions.

John Stuart Mill's Ethics

The ethical theory of John Stuart Mill (1806-1873) is most extensively articulated in his classical text *Utilitarianism* (1861). Its goal is to justify the utilitarian principle as the foundation of morals. This principle says actions are right in proportion as they tend to promote overall human happiness. So, Mill focuses on consequences of actions and not on rights nor ethical sentiments.

This article primarily examines the central ideas of his text *Utilitarianism*, but the article's last two sections are devoted to Mill's views on the freedom of the will and the justification of punishment, which are found in *System of Logic* (1843) and *Examination of Sir William Hamilton's Philosophy* (1865), respectively.

Educated by his father James Mill who was a close friend to [Jeremy Bentham](#), John Stuart Mill came in contact with utilitarian thought at a very early stage of his life. In his *Autobiography* he claims to have introduced the word "utilitarian" into the English language when he was sixteen. Mill remained a utilitarian throughout his life. Beginning in the 1830s he became increasingly critical of what he calls Bentham's "theory of human nature". The two articles "Remarks on Bentham's Philosophy" (1833) and "Bentham" (1838) are his first important contributions to the development of utilitarian thought. Mill rejects Bentham's view that humans are unrelentingly driven by narrow self-interest. He believed that a "desire of perfection" and sympathy for fellow human beings belong to human nature. One of the central tenets of Mill's political outlook is that, not only the rules of society, but also people themselves are capable of improvement.

2. Mill's Theory of Value and the Principle of Utility

Mill defines "utilitarianism" as the creed that considers a particular "theory of life" as the "foundation of morals" (CW 10, 210). His view of theory of life was monistic: There is one thing, and one thing only, that is intrinsically desirable, namely pleasure. In contrast to a form of hedonism that conceives pleasure as a homogeneous matter, Mill was convinced that some types of pleasure are more valuable than others in virtue of their inherent qualities. For this reason, his position is often called "qualitative hedonism". Many philosophers hold that qualitative hedonism is no consistent position. Hedonism asserts that pleasure is the only intrinsic value. Under this assumption, the critics argue, there can be no evaluative basis for the distinction between higher and lower pleasures. Probably the first ones to raise this common objection were the British idealists F. H. Bradley (1876/1988) and T. H. Green (1883/2003).

Which inherent qualities make one kind of pleasure better than another, according to Mill? He declares that the more valuable pleasures are those which employ "higher faculties" (CW 10, 211). The list of such better enjoyments includes "the pleasures of intellect, of the feelings and imagination, and of the moral sentiments" (CW 10, 211). These enjoyments make use of highly developed capacities, like judgment and empathy. In one of his most famous sentences, Mill affirms that it "is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied" (CW 10, 212). This seems to be a surprising thing to say for a hedonist. However, Mill thought that we have a solid empirical basis for this view. According to him, the best obtainable evidence for value claims consists in what all or almost all people judge as valuable across a vast variety of cases and cultures. He makes the empirical assertion that all or almost all people prefer a "manner of existence" (CW 10, 211) that employs higher faculties to a manner of existence which does not. The fact that "all or almost all" who are acquainted with pleasures that employ higher faculties agree that they are preferable to the lower ones, is empirical evidence for the claim that they are indeed of higher value. Accordingly, the best human life ("manner of existence") is one in which the higher faculties play an adequate part. This partly explains why he put such great emphasis on education.

3. Morality as a System of Social Rules

The fifth and final chapter of *Utilitarianism* is of unusual importance for Mill's theory of moral obligation. Until the 1970s, the significance of the chapter had been largely overlooked. It then became one of the bridgeheads of a revisionist interpretation of Mill, which is associated with the work of David Lyons, John Skorupski and others.

Mill worked very hard to hammer the fifth chapter into shape and his success has great meaning for him. Towards the end of the book he maintains the “considerations which have now been adduced resolve, I conceive, the only real difficulty in the utilitarian theory of morals.” (CW 10, 259)

At the beginning of *Utilitarianism*, Mill postulates that moral judgments presume rules (CW 10, 206). In contrast to Kant who grounds his ethical theory on *self-imposed* rules, so-called maxims, Mill thinks that morality builds on *social* rules. But what makes *social* rules *moral* rules? Mill’s answer is based on a thesis about how competent speakers use the phrase “morally right” or “morally wrong”. He maintains that we name a type of action morally wrong if we think that it should be sanctioned either through formal punishment, public disapproval (external sanctions) or through a bad conscience (internal sanctions). This is the critical difference between “morality and simple expediency” (CW 10, 246). Wrong or inexpedient actions are those that we cannot recommend to a person, like harming oneself. But in contrast to immoral actions, inexpedient actions are not worthy of being sanctioned.

Mill differentiates various spheres of action. In his *System of Logic* he names morality, prudence and aesthetics as the three departments of the “Art of Life” (CW 8, 949). The principle of utility governs not only morality, but also prudence and taste (CW 8, 951). It is not a moral principle but a meta-principle of practical reason (Skorupski 1989, 310-313).

There is a field of action in which moral rules obtain, and a “person may rightfully be compelled to fulfill” them (CW 10, 246). But there are also fields of action, in which sanctions for wrong behavior would be inappropriate. One of them is the sphere of self-regarding acts with which Mill deals in *On Liberty*. In this private sphere we can act at our convenience and indulge in inexpedient and utterly useless behavior as long as we do not harm others.

It is fundamental to keep in mind that Mill looks into morality as a social practice and not as autonomous self-determination by reason, like Kant. For Kantians, moral deliberation determines those actions which we have the most reason to perform. Mill disagrees; for him, it makes sense to say that “A is the right thing to do for Jeremy, but Jeremy is not morally obliged to do A.” For instance, even if Jeremy is capable of writing a brilliant book that would improve the life of millions (and deteriorate none), he is not morally obliged to do so. According to Mill, our moral obligations result from the *justified* part of the moral code of our society; and the task of moral philosophy consists in bringing the moral code of a society in better accordance with the principle of utility.

6. Applying the Standard of Morality

In “Whewell on Moral Philosophy” (1852), Mill rejects an objection raised by one of his most competent philosophical adversaries. Whewell claimed that utilitarianism permits murder and other crimes in particular circumstances and is therefore incompatible with our considered moral judgments. Mill’s discussion of Whewell’s criticism is exceedingly helpful in clarifying his ethical approach:

Take, for example, the case of murder. There are many persons to kill whom would be to remove men who are a cause of no good to any human being, of cruel physical and moral suffering to several, and whose whole influence tends to increase the mass of unhappiness and vice. Were such a man to be assassinated, the balance of traceable consequences would be greatly in favour of the act. (CW 10, 181)

Mill gives no concrete case. Since he wrote – together with his wife Harriet Taylor – a couple of articles on horrible cases of domestic violence in the early 1850s, he might have had the likes of Robert Curtis Bird in mind, a man who tortured his servant Mary Ann Parsons to death [see CW 25 (The Case of Mary Ann Parsons), 1151-1153]. Does utilitarianism require us to kill such people who are the “cause of no good to any human being, of cruel physical and moral suffering to several”? Mill answers in the negative. His main point is that nobody’s life would be safe if people were allowed to kill others whom they believe to be a source of unhappiness (CW 10, 182). Thus, a general rule that would allow to “remove men who are a cause of no good” would be worse than a general rule that does not allow such acts. People should follow the rule not to kill other humans

because the general observance of this rule tends to promote the happiness of all.

This argument can be interpreted in a rule utilitarian or an indirect act utilitarian fashion. Along indirect act utilitarian lines, one could maintain that we would be cognitively overwhelmed by the task of calculating the consequences of any action. We therefore need rules as touchstones that point us to the path of action which tends to promote the greatest general happiness. Mill compares, in a critical passage, the core principles of our established morality (which he also calls “secondary principles”) with the *Nautical Almanack*, a companion for navigating a voyage (CW 10, 225). Just as the *Nautical Almanack* is not first calculated at sea, but instead exists as already calculated, the agent must not in individual cases calculate the expected utility. In his moral deliberation the agent can appeal to secondary principles, such as the prohibition of homicide, as an approximate solution for the estimated problem.

Apparently, the act utilitarian interpretation finds further support in a letter Mill wrote to John Venn in 1872. He states:

I agree with you that the right way of testing actions by their consequences, is to test them by their natural consequences of the particular actions, and not by those which would follow if everyone did the same. But, for the most part, considerations of what would happen if everyone did the same, is the only means we have of discovering the tendency of the act in the particular case. (CW 17, 1881)

Mill argues that in many cases we can assess the *actual, expected* consequences of an action, only if we hypothetically consider that all would act in the same manner. This means we *recognize* that the consequences of *this particular* action would be damaging if everyone acted that way. A similar consideration is found in the Whewell essay. Here Mill argues: If a hundred breaches of rule (homicides, in this case) led to a particular harm (murderous chaos), then a single breach of rule is responsible for a hundredth of the harm. This hundredth of harm offsets the expected utility of this particular breach of rule (CW 10, 182). Mill believes that the breach of the rule is wrong because it is actually harmful. The argument is questionable because Mill overturns the presumption he introduces: that the actual consequences of the considered action would be beneficial. If the breach of the rule is actually harmful, then it is to be rejected in every conceivable version of utilitarianism. The result is trivial then and misses the criticism that act utilitarianism has counter-intuitive implications in particular circumstances.

There is one crucial difficulty with the interpretation of Mill as an indirect act utilitarian regarding moral obligation. If the function of rules was in fact *only* epistemic, as suggested by indirect act utilitarianism, one would expect that the principle of utility – when the epistemic conditions are satisfactory – can be and should be directly applied. But Mill is quite explicit here. The utilitarian principle should only be applied when moral rules conflict: “We must remember that only in these cases of conflict between secondary principles is it requisite that first principles should be appealed to.” (CW 10, 226). From an act utilitarian view regarding moral obligation, this is implausible. Why should one be morally obliged to follow a rule of which one positively knows that its observance in a particular case will not promote general utility?

Coming back to the example, it is important to remember that “the balance of traceable consequences would be greatly in favour of the act [of homicide].” (CW 10, 181) Thus, according to an act utilitarian approach regarding moral obligation it would be morally allowed, if not required, to kill the man.

As mentioned, Mill arrives at a different conclusion. His position can be best understood with recourse to the distinction between the *theory of objective rightness* and the *theory of moral obligation* introduced in the last section. Seen from the perspective of an all-knowing and impartial observer, it is – in regard to the given description – objectively right to perpetrate the homicide. However, moral laws, permissions, and prohibitions are not made for omniscient and impartial observers, but instead for cognitively limited and partial beings like humans whose actions are mainly guided by acquired dispositions. Their capacity to recognize what would be objectively right is imperfect; and their ability to motivate themselves to do the right thing is limited. As quoted before in his “Remarks on Bentham’s Philosophy” (1833), he states that some violations of the established moral code are simply unthinkable for the members of society:

people recoil “from the very thought of committing” (CW 10, 12) particular acts. Because humans cannot reliably recognize objective rightness and, in critical cases, cannot bring themselves to act objectively right, they are not obliged to maximize happiness. For ought implies can. In regard to the given description, the fact that the assassination of a human would be objectively right does not imply that the assassination of this human would be morally imperative or allowed. In other words: Mill differentiates between the *objectively right act* and the *morally right act*. With this he can argue that the assassination would be forbidden (*theory of moral obligation*). To enact a forbidden action is *morally wrong*. As noted, Mill’s theory allows for the possibility that an action is objectively right, but morally wrong (prohibited). An action *can* be wrong (bearing unhappiness), but its enactment would be no less morally right (Lyons 1978/1994, 70).

Thus, Mill’s considered position should be interpreted in the following way: First, the objective rightness of an act depends upon actual consequences; second, in order to know what we are morally obliged to do we have to draw on justified rules of the established moral code.

7. The Meaning of the First Formula

What has been said about Mill’s conception of morality as a system of social rules is relevant for the interpretation of Mill’s *First Formula* of utilitarianism. The *Formula* says that actions are right “in proportion as they tend to promote happiness” and wrong “as they tend to produce the reverse of happiness” (CW 10, 210). Roughly said, actions are right insofar as they facilitate happiness, and wrong insofar as they result in suffering. Mill does not write “morally right” or “morally wrong”, but simply “right” and “wrong”. This is important. Mill emphasizes in many places that virtuous actions can exhibit a negative balance of happiness in a singular case. If the word “moral” occurred in the *First Formula*, then the noted virtuous actions would be, for Mill, *morally wrong*. But as we have seen, this is not his view. Virtuous actions are morally right, even if they are objectively wrong under particular circumstances.

Accordingly, the *First Formula* is not to be interpreted as drafting a *moral duty*. It is a general statement about what makes actions right (reasonable, expedient) or wrong. The *First Formula* gives a general characterization of practical reason. It says that the promotion of happiness makes an action objectively right (but not necessarily morally right); or, as Mill says in his *System of Logic*, “the promotion of happiness is the ultimate principle of Teleology” (CW 8, 951). An action is *objectively* right if it maximizes happiness; however, an action is *morally* right if it is in accordance with social rules which are protected by internal and external sanctions and which tend to promote general utility. Subsets of right ones are morally right actions; subsets of wrong actions are morally wrong.

Mill’s differentiation between a moral and a non-moral sphere of action is not far from our everyday understanding. We generally believe that not all actions must be judged in regard to a moral point of view. This does not exclude us from valuing actions, which are not in the moral realm, in regard to prudence. Less evident is how one should take Mill’s claim that the promotion of happiness can be understood as a general principle of rightness even with respect to artistic production. Many artists would presumably not be comfortable with the thesis that good art arises from the goal of facilitating the happiness of humankind. This however is not what Mill means. Apart from cases of conflict between secondary principles, the *First Formula* does not guide action. Just as Mill speaks in a moral context about how noble characters will not strive to maximize general happiness (CW 8, 952), he could argue in an aesthetic context that artists should work from a purely aesthetic point of view. The *rules* of artistic judgments, nonetheless, are justified through their contribution to the flourishing of human life.

To summarize the essential points: Mill can be characterized as an act utilitarian in regard to the theory of objective rightness, but as a rule utilitarian in regard to the theory of moral obligation. He defines morality as a system of rules that is protected by sanctions. The principle of utility is not a part of this system, but its fundamental justification (the “foundation of morality” (CW 10, 205)).

8. Right in Proportion and Tendencies

(i) For contemporary readers it is striking that Mill’s *First Formula* does not explicitly relate to

maximization. Mill does not write, as one might expect, that only the action which leads to the best consequences is right. In other places in the text we hear of the “promotion” or “multiplication” of happiness, and not of the “maximization”. Alone does the “Greatest Happiness Principle” explicitly refer to maximization. The actual formula, in contrast, has to do with gradual differences (right in proportion). Actions which add to the sum of happiness in the world but fail to maximize happiness thus can be right, even if to a lesser degree.

This is confusing insofar as it would be unreasonable to prefer that which is worse to that which is better. For every good there is a better that one should reasonably choose until one succeeds to the best. If the *First Formula* expresses the ideal of practical reason, then one should expect that it requires maximization. Maybe Mill’s point is that the search for a global best option would exceed the cognitive capabilities of humans. He probably does not want to suggest that an agent should not choose the best local option. But the local best option must not represent the objective (global) best. This may be the reason why Mill does not refer to maximization in the formula of utility.

(ii) A further complication arises with the word “tend”. According to the formula of utility, actions are *more or less* correct insofar as they facilitate happiness (CW 10, 210). It is doubtlessly not the same to say that an action is right if it *actually* facilitates happiness, or to say that it is right if it *tends* to facilitate happiness. The model seems to be roughly this: At the neutral point of the preference scale, actions have the tendency – in regard to the status quo – to neither increase nor decrease the mass of utility in the world. All actions that tend to facilitate happiness are right, all actions that tend to be harmful are wrong, but all are not in the same measure. An action has a high positive value on the scale of preference, if its tendency to facilitate happiness is high. An action has highly negative value on the preference scale, if its tendency to evoke unhappiness is high. But what does the concept “tendency” mean precisely?

In everyday language, we often use the word “tend” in the sense of “will probably lead to”. That an action tends to produce a particular consequence means that this consequence has a high probability. Mill could have wanted to say that an action is right in proportion to the probability with which it promotes happiness. This makes sense when we compare options that produce the same amount of happiness. But what about cases in which two actions produce different amounts of pleasure? One plausible answer is that both dimensions must be regarded: the amount of happiness *and* the probability of its occurrence. Action *A* is better than action *B*, if the expected happiness for *A* is greater than the expected happiness for *B*. If one reads Mill this way, then “in proportion” relates to “promote” *and* to “tend”. The best action is one that maximizes the amount of expected happiness.

9. Utility and Justice

In the final chapter of *Utilitarianism*, Mill turns to the sentiment of justice. Actions that are perceived as unjust provoke outrage. The spontaneity of this feeling and its intensity makes it impossible for it to be ignored by the theory of morals. Mill considers two possible interpretations of the source of the sentiment of justice: first of all, that we are equipped with a sense of justice which is an independent source of moral judgment; second, that there is a general and independent principle of justice. Both interpretations are irreconcilable with Mill’s position, and thus it is no wonder that he takes this issue to be of exceptional importance. He names the integration of justice the only real difficulty for utilitarian theory (CW 10, 259).

Mill splits this problem of integration into three tasks: The first consists in explaining the intensity and spontaneity of the sentiment of justice. The second task is to make plausible that the various types of judgments about justice can be traced back to a systematic core; and the third task consists in showing that the principle of utility constructs this core.

In a nutshell, Mill explains the sentiment of justice as the sublimation of the impulse to take revenge for perceived mortifications of all kinds. Mill sees vengeance as “an animal desire” (CW 10, 250) that operates in the service of self-preservation. If it is known that one will not accept interventions in spheres of influence and interest, the probability of such interventions dwindles. The preparedness to take revenge tends to deter aggression in the first place. Thus, a reputation for vindictiveness – at first glance an irrational trait – arguably has survival value. This helps to

explain why the sentiment is so widespread and vehement.

Our sentiment of justice, for Mill, is based on a refinement and sublimation of this animal desire. Humans are capable of empathizing such that the pleasure of others can instill one's own pleasure, and the mere sight of suffering can cause own suffering. The hurting of another person or even an animal may therefore produce a very similar affect as the hurting of one's own person. Mill considers the extension of the animal impulse of vengeance on those with whom we have sympathy as "natural" (CW 10, 248), because the social feelings are for him natural. This natural extension of the impulse of revenge with the help of the social feelings represents a step in the direction of cultivating and refining human motivation. People begin to feel outrage when the interests of the members of their tribe are being violated or when shared social rules are being disregarded.

Gradually, sympathy becomes more inclusive. Humans discover that co-operation with people outside the tribe is advantageous. The "human capacity of enlarged sympathy" follows suit (CW 10, 248).

As soon as humans begin to think about which parts of the moral code of a society are justified and which parts are not, they inevitably begin to consider consequences. This often occurs in non-systematic, prejudiced or distorted ways. Across historical periods of times, the correct ideas of intrinsic good and moral rightness will gradually gain more influence. Judgments about justice approximate progressively the requirements of utilitarianism: The rules upon which the judgments about justice rest will be assessed in light of their tendency to promote happiness. To summarize: Our sentiment of justice receives its intensity from the "animal desire to repel or retaliate a hurt or damage to oneself", and its morality from the "human capacity of enlarged sympathy" and intelligent self-interest (CW 10, 250).

According to Mill, when we see a social practice or a type of action as unjust, we see that the *moral rights* of persons were harmed. The thought of moral rights is the *systematic core* of our judgments of justice. Rights breed perfect obligations, says Mill. Moral rights are concerned with the basic conditions of a good life. They protect an "extraordinarily important and impressive kind of utility." (CW 10, 250-251). Mill subsumes this important and impressive kind of utility under the term security, "the most vital of all interests" (CW 10, 251). It comprises such things as protection from aggression or starvation, the possibility to shape one's own life unmolested by others and enforcement of contracts. Thus, the requirements of justice "stand higher in the scale of social utility" (CW 259). To have a moral right means to have something that society is morally required to guard either through the compulsion of law, education or the pressure of public opinion (CW 10, 250). Because everyone has an interest in the security of these conditions, it is desirable that the members of society reciprocally guarantee each other "to join in making safe for us the very groundwork of our existence" (CW 10, 251). Insofar as moral rights secure the basis of our existence, they serve our natural interest in self-preservation – this is the reason why their harm calls forth such intense emotional reactions. The interplay of social feelings and moral education explains, in turn, why we are not only upset by injustices when we personally suffer, but also when the elemental rights of *others* are harmed. This motivates us to sanction the suffering of others as unjust. Moral rights thus form the "most sacred and binding part of all morality" (CW 10, 255). But they do not exhaust the moral realm. There are imperfect obligations which have no correlative right (CW 10, 247).

The thesis that moral rights form the systematic core of our judgments of justice is by no means unique to utilitarianism. Many people take it to be evident that individuals have absolute, inalienable rights; but they doubt that these rights can be grounded in the principle of utility. Intuitionists may claim that we recognize moral rights spontaneously, that we have intuitive knowledge of them. In order to reject such a view, Mill points out that our judgments of justice do not form a systematic order. If we had a sense of justice that would allow us to recognize what is just, similar to how touch reveals forms or sight reveals color, then we would expect that our corresponding judgments would exhibit a high degree of reliability, definitude and unanimity. But experience teaches us that our judgments regarding just punishments, just tax laws or just remuneration for waged labor are anything but unanimous. The intuitionists must therefore mobilize a *first principle* that is independent of experience and that secures the unity and

consistency of our theory of justice. So far they have not succeeded. Mill sees no suggestion that is plausible or which has been met with general acceptance.

10. The Proof of Utilitarianism

What Mill names the “proof” of utilitarianism belongs presumably to the most frequently attacked text passages in the history of philosophy. Geoffrey Sayre-McCord once remarked that Mill seems to answer by example the question of how many serious mistakes a brilliant philosopher can make within a brief paragraph (Sayre-McCord 2001, 330). Meanwhile the secondary literature has made it clear that Mill’s proof contains no logical fallacies and is less foolish than often portrayed.

It is found in the fourth part, “Of What Sort of Proof the Principle of Utility is Susceptible”, of *Utilitarianism*. For the assessment of the proof two introductory comments are helpful. Already at the beginning of *Utilitarianism*, Mill points out that “questions of ultimate ends are not amenable to direct proof.” (CW 10, 207). Notwithstanding, it is possible to give reasons for theories about the good, and these considerations are “equivalent to proof” (CW 10, 208). These reasons are empirical and touch upon the careful observation of oneself and others. More cannot be done and should not be expected in a proof re ultimate ends.

A further introductory comment concerns the basis of observation through which Mill seeks to support utilitarianism. In moral philosophy the appeal to intuitions plays a prominent role. They are used to justify moral claims and to check the plausibility of moral theories. The task of thought-experiments in testing ethical theories is analogous to the observation of facts in testing empirical theories. This suggests that intuitions are the right observational basis for the justification of first moral principles. Mill, however, was a fervent critic of intuitionism throughout his philosophical work. In his *Autobiography* he calls intuitionism “the great intellectual support of false doctrines and bad institutions.” (CW 1, 232). Mill considered the idea that truths can be known *a priori*, independently of observation and experience, to be a stronghold of conservatism.

His argument against intuitionistic approaches to moral philosophy has two parts. The first part points out that intuitionists have not been able to bring our intuitive moral judgments into a system. There is neither a complete list of intuitive moral precepts nor a basic principle of morality which would found such a list (CW 10, 206).

The second part of the Millian argument consists in an explanation of this result: What some call moral intuition is actually the result of our education and present social discourse. Society inculcates us with our moral views, and we come to believe strongly in their unquestionable truth. There is no system, no basic principle in the moral views of the Victorian era though. In *The Subjection of Women*, Mill caustically criticizes the moral intuitions of his contemporaries regarding the role of women. He finds them incompatible with the basic principles of the modern world, such as equality and liberty. Because the first principle of morality is missing, intuitionist ethics is in many regards just a decoration of the moral prejudices with which one is brought up – “(...) not so much a guide as a consecration of men’s actual sentiments” (CW 10, 207).

What we need, Mill contends, is a basis of observation that verifies a first principle, a principle that is capable of bringing our practice of moral judgments into order. This elemental observational basis – and this is the core idea in Mill’s proof – is human aspiration.

His argument for the utilitarian principle – if not a deductive argument, an argument all the same – involves three steps. First, Mill argues that it is reasonable for humans to aspire to one’s own well-being; second, that it is reasonable to support the well-being of all persons (instead of only one’s own); and third, that well-being represents the *only* ultimate goal and the rightness of our actions is to be measured exclusively in regard to the balance of happiness to which they lead (CW 10, 234).

Let us turn to the *first step of the argument*. Upon an initial reading it seems in fact to have little success. Mill argues that one’s own well-being is *worthy* of striving for because each of us strives for his or her own well-being. Here he leans on a questionable analogy: “The only proof capable of being given that an object is visible, is that people actually see it. [...] In like manner, I

apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it.” (CW 10, 234).

Can a more evident logical fallacy be given than the claim that something is *worthy* of striving for because it is factually sought? But Mill in no way believes that the relation between desirable and desired is a matter of definition. He is not saying that desirable objects are by definition objects which people desire; he writes instead that what people desire is the only *evidence* for what is desirable. If we want to know what is ultimately desirable for humans, we have to acquire observational knowledge about what humans ultimately strive for.

Mill’s argument is simple: We know by observation that people desire their own happiness. With a conclusion that Mill calls “inductive”, and to which he ascribes a central role in regard to our acquisition of knowledge, we succeed to the general thesis that all humans finally aspire to their happiness. This inductive conclusion serves as evidence for the claim that one’s own happiness is not only desired, but *desirable*, worthy of aspiration. Mill thus supports the thesis that one’s own happiness is an ultimate good to oneself with the observation that every human ultimately strives for his or her own well-being.

On this basis, Mill concludes in the *second step* of his proof that the happiness of all is also a good: “...each person’s happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of all persons.” (CW 10, 234).

The “therefore” in the cited sentence above has evoked many a raised eyebrows. Does Mill claim here that each person tries to promote the happiness of all? This seems to be patently wrong. In a famous letter to a Henry Jones, he clarifies that he did not mean that every person, in fact, strives for the general good. “I merely meant in this particular sentence to argue that since *A*’s happiness is a good, *B*’s a good, *C*’s a good, &c., the sum of all these goods must be a good.” (CW 16, 1414, Letter 1257).

Indeed, in the “particular sentence” he just concludes that general happiness is a “good to the aggregate of all persons.” Nonetheless, one may doubt that Mill adequately responds to Jones’ reservations. It is unclear what it means that general happiness is the good of the aggregate of all persons. Neither each person, nor the aggregate of all persons seem to strive for the happiness of all. But Mill’s point in the second step of the argument is arguably a more modest one.

He simply wanted to vindicate the claim that if each person’s happiness is a good to each person, then we are entitled to conclude that general happiness is also a good. As he says in the letter to Jones: “the sum of all these goods must be a good.” Similar to the first step of the argument we have here an epistemic relationship: The fact that each person is striving for his or her own happiness is evidence that happiness as such (regardless to whom) is valuable. If happiness as such is valuable, it is not unreasonable to promote the well-being of all sentient beings. With this, the second step of the argument is complete. The result may seem meager at first. That it is not unreasonable to promote the happiness of all appears to be no particularly controversial claim. On closer inspection, however, Mill’s conclusion is quite interesting since it imposes pressure on self-interest theories of practical rationality. The “notion that self-interest possesses a special, underived rationality (...) seems suddenly to require justification.” (Skorupski 1989, 311). What Mill fails to show is that each person has most reason to promote the general good. One should note, however, that the aim of the proof is not to answer the question why one should be moral. Mill does not want to demonstrate that we have reason to prefer general happiness to personal happiness.

Hedonism states not only that happiness is intrinsically good, but also that it is *the only* good and thus the only measure for our action. To show this, is the goal of the *third step* of the proof. Mill’s reflections in this step are based on psychological hedonism and the principle of association. According to Mill, humans cannot desire anything except that which is either an instrument to or a component of happiness. He concedes that people seem to strive for every possible thing as ultimate ends. Philosophers may pursue knowledge as their ultimate goal; others value virtue, fame or wealth. Corresponding to his basic thesis that “the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it” (CW 10, 234), Mill must consider the possibility that knowledge, fame or wealth have intrinsic value.

He blocks this inference with the thesis that humans do not “naturally and originally” (CW 10, 235) desire other goods than happiness. That knowledge, virtue, wealth or fame is seen as intrinsically valuable is due to the operation of the principle of association. In the course of our socialization, goods, like knowledge, virtue, wealth or fame acquire value by their association with pleasure. A philosopher came to experience knowledge as pleasurable, and this is why he desires it. Humans strive for virtue and other goods only if they are associated with the natural and original tendency to seek pleasure and avoid pain. Virtue, knowledge or wealth can thus become parts of happiness. At this point, Mill declares that the proof is completed.

11. Evaluating Consequences

According to Mill’s *Second Formula* of the utilitarian standard, a good human life must be rich in enjoyments, in both quantitative and qualitative respects. A manner of existence without access to the higher pleasures is not desirable: “It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied.” (CW 10, 212).

The life of Socrates is better because no person who is familiar with higher pleasures will trade the joy of philosophizing against an even infinite amount of lower pleasures, Mill suggests. This does not amount to a modern version of Aristotle’s view that only a life completely devoted to theoretical activity is desirable. One must not forget that Mill is a hedonist after all. What kind of life is joyful and therefore good for a particular person depends upon many factors, such as tastes, talents and character. There are a great variety of lifestyles that are equally good. But Mill insists that a human life that is completely deprived of higher pleasures is not as good as it could be. It is not a desirable “mode of existence”, nothing a “competent judge” would choose.

Utilitarianism demands that we establish and observe a system of social, legal and moral rules that enables all mankind to have the best life possible, a life that is “as rich as possible in enjoyments, both in point of quantity and quality” (CW 10, 214). Mill’s statement that every human has an equal claim “to all the means of happiness” (CW 10, 258), belongs in this context. Society must make sure that the social-economic preconditions of a non-impooverished life prevail. In one text passage, Mill even includes the happiness of animals. Animals, too, should have the best possible life, “so far as the nature of things admits” (CW 10, 214).

The *Second Formula* maintains that a set of social rules *A* is better than the set *B*, if in *A* less humans suffer from an impooverished, unhappy life and more enjoy a fulfilled, rich life than in *B*.

More difficult is the question how to evaluate scenarios that involve unequal population sizes. With Mill there is no explicit unpacking of this problem; but his advocacy of the regulation of birth gives us at least an indication of the direction in which his considerations would go. Let us consider the following example: Which world would be better: world *X* in which 1000 humans have a fulfilled life and 100 a bad one, or world *Y* in which 10000 humans have a fulfilled life and 800 an impooverished one? The answer to this question depends on whether we focus on the minimizing the number of bad lives or on maximizing the number of good lives, and whether we measure this absolutely or relatively to the total population.

(i) One possible answer concerns the minimization of the number of *bad lives*. This can mean the absolute number of humans with joyless or impooverished lives. If one answers this way, then world *X* would be better than world *Y* because in this world the absolute number of humans with bad lives would be less. But it is also possible to think of the *Second Formula* as a statement about the relative number of humans with bad lives; in this case world *Y* would be preferable.

(ii) Another possible answer emphasizes the maximization of fulfilled lives. If one follows this interpretation, then world *Y* is better than world *X* because in this world absolute and relative measurements suggest that more humans have fulfilled lives.

Under the influence of Malthus, Mill insisted throughout his work that the problem of poverty is to be resolved only through a reduction of the population number – as noted, he encouraged the regulation of birth. This proposal is reconcilable with all three interpretations, but does not bear any relation to the question concerning which of the interpretations he could have preferred. One can speculate how Mill would answer, but there is not clear textual basis.

A further theme that Mill does not address concerns the problem of measurement and the interpersonal comparison of quantities of happiness. From an utilitarian point of view, other things being equal, it makes no moral difference whether *A* or *B* experiences an equal quantity of happiness (CW 10, 258). A quantity of happiness for *A* bears precisely as much value as a quantity of happiness for *B*. But this answers neither the question of measurement nor the question of the comparison of interpersonal utility. Can quantities of happiness be measured like temperatures? The philosopher and economist Francis Edgeworth spoke in his 1881 *Mathematical Psychics* of a fictitious instrument of measurement, a hedonimeter, with whose help the quantities of pleasure and pain could be determined with scientific accuracy.

Or do amounts of happiness have to be assessed approximately, such that Harriet Taylor for example can say that she is happier today than she was yesterday. Interpersonal comparisons of utility are confronted with the related question whether and under which conditions one can say that, for instance, Harriet Taylor and John Stuart Mill experience an equal amount of happiness.

Mill gave both themes little attention. But probably he was convinced that precise measurement and comparison of interpersonal utility would not be needed, maybe not even possible. One often does not need a thermometer to discern whether or not an object is warmer than another. Similarly, in many cases we do not need something like a hedonimeter to judge whether the condition of world *A* is better than that of world *B*. We need only a reasonable degree of experience and the capacity to empathize. Often, though, we may be unsure what to say. Which of two systems of income tax, for instance, promotes general happiness more? Mill's position here seems to be that we have to decide questions like these by means of public debate and not by means of a hedonimeter.

Regarding moral rights, "the most sacred and binding part of all morality" (CW 10, 255), all competent judges seem to agree that they promote general happiness. Our capacity to estimate quantities and qualities of happiness is thus sufficiently good in order to conclude that a society that does protect "the most vital of all interests" (CW 10, 251) is better than a society that does not.