

JOHN STUART MILL'S FAILED THOUGHT EXPERIMENT

CLAUDIU BACIU

Institutul de Filosofie și Psihologie „Constantin Rădulescu-Motru”
al Academiei Române

Abstract: Mill's utilitarian ethics is based on two assumptions: first, the Epicurean view that man is constantly in search of happiness; second, a thought experiment in which all human beings would gather together and share their life experiences. As a result of such sharing, they could understand that a significant majority of them prefer certain experiences that, as a consequence, will be seen as a ground for future universal human values. My presentation challenges the idea that universal values can emerge statistically through a consensus of individuals who lack any previous education (i.e., a system of beliefs that shape 'a priori' their experiences) and who thus can live wholly isolated from existing values.

Keywords: Mill, Kant, Thought Experiment, Contradiction, Competent Judges, Highest Good

1. MILL'S CRITIQUE OF KANT

Although it might not be evident at first sight, Mill's utilitarianism has many affinities with Kantian ethics. In fact, one could say that this form of utilitarianism is an attempt to accommodate the latter with the historicity of man as a fundamental principle and a discovery of 19th-century philosophy and science. In this respect, Mill's idea of the "competent judge" echoes Kant's concept of the rational person, attempting to connect the principles of morality with an imagined end of (pre)history from where, thereafter, they could be raised back again to the level of a sort of transcendental condition of morality. Because on the one hand he shares Epicurus' view that the most fundamental desire of the human soul is his striving toward happiness, and happiness, on the other hand, is a concept that remained so unclear for millennia since it entered the philosophical discourse, Mill imagines a thought experiment in which at a particular moment in history (most probably at the end of it) those individuals who managed to savor all possible pleasures reach an agreement that there is a hierarchy of pleasures and that the higher or nobler or spiritual pleasures are always preferable to the lower, coarser and bodily ones. Mill expresses this idea as follows:

“If I am asked what I mean by difference of quality in pleasures, or what makes one pleasure more valuable than another, merely as a pleasure, except its being greater in amount, there is but one possible answer. Of two pleasures, if

there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure. If one of the two is, by those who are competently acquainted with both, placed so far above the other that they prefer it, even though knowing it to be attended with a greater amount of discontent, and would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a superior quality, so far outweighing quantity as to render it, in comparison, of small account.”¹

From the idea that there could exist such “competent judges” then, he deduces that what such persons would consider to be the higher pleasures has to become a fundamental benchmark for all human behavior and thus receive a relevant moral meaning. In other words, because the “competent judges” have understood that higher pleasures are the ones that come the closest to the much sought-after concept of happiness, the whole of humanity – which, alas, did not manage to experience such a wide variety of pleasures – must trust them and give up immediate inclinations and impulses in order to reach those superior joys.

The concept of “competent judges” seems to be close to the Kantian concept of a rational person. Both of them are idealized versions of the concrete human being, something that cannot be reached in any real historical development but that nonetheless is taken as a benchmark for moral assessment. However, there is a fundamental difference between them. As abstract as the Kantian concept of the rational person is, it has a feature that can be concretely discovered, namely rationality, which consists of the capacity to create universal and necessary knowledge, as we can witness it in Newtonian science, mathematics, and logic. These sciences, based on those two fundamental features, were facts for Kant; they were well-established and recognized forms of knowledge² at the moment when Kant developed his moral philosophy and could be taken, therefore, as a real ‘groundwork’ for his ethics. In comparison, the idea of spontaneously reaching a hierarchy of pleasures is a utopian picture that can never be given since human history is always “in the making”. What is more, as we will see below, there is not only a temporal impossibility but also a logical impossibility.

The two concepts that ground Kantian and Mill’s ethics are very different in nature. Whereas for the Kantian rational person, although the latter seems to have rationality as its very nature, in fact, that character must always be proved or actualized. You cannot have universality and necessity in themselves as given properties of a rational person; they are rather tendencies that actualize in the form of something that has a universal and necessary character. They are similar to the talent of a painter which cannot be shown as a quality of that painter until it has been actualized through his paintings. In comparison, in the case of the ‘competent judges,’ we deal with beings that actually possess the knowledge of all possible pleasures, who, therefore, effectively own the knowledge of those pleasures already

¹ John Stuart Mill, *Utilitarianism*, Chicago, The University of Chicago Press, 1902, p. 16.

² Imm. Kant, *Critique of Pure Reason*, translated and edited by Paul Guyer and Allen W. Wood, Cambridge, Cambridge University Press, p. 108–109.

as their attributes. This conceptual difference seems to have been overlooked by Mill, with important consequences marking his critique of Kantian ethics.

Mill expects Kant to show that violating moral duties should contradict the ultimate principle of his ethics. This is because he considers that, as happens in general in sciences, the content of Kantian ethics, as a science of morality, should be derived from its principles, and, thus, it must never contradict those principles. In this vein, according to Mill, the moral duties presented by Kant should be shown as not contradicting the initial principle of his ethics, which is the categorical imperative.

“This remarkable man, writes Mill, whose system of thought will long remain one of the landmarks in the history of philosophical speculation, does, in the treatise in question, lay down a universal first principle as the origin and ground of moral obligation; it is this: ‘so act that the rule on which thou actest would admit of being adopted as a law by all rational beings.’ But when he begins to deduce from this precept any of the actual duties of morality, he fails, almost grotesquely, to show that there would be any contradiction, any logical (not to say physical) impossibility, in the adoption by all rational beings of the most outrageously immoral rules of conduct. All he shows is that the consequences of their universal adoption would be such as no one would choose to incur”³.

Mill overlooks Kant’s general approach, which is not that of an ordinary science but of transcendental philosophy. In this respect, Kant is interested in explaining the *possibility* of morality, i.e., of concretely existing morality, and not in grounding ethics as a common science. Indeed, he offers a general criterion for any moral duty. But duty is not something that thereafter is somehow *deduced* from that criterion, but it is examined to see if it is compatible with the criterion or not. And, in this respect, of course, Kant can offer nothing else than “consequences”, i.e., exactly what Mill criticizes in Kant.

A logical contradiction exists between two statements of which one claims what the other denies. For example, the propositions “Fire burns the wood” and “Fire doesn’t burn the wood” are contradictory propositions, one stating what the other negates. It is obvious that no maxim of any duty could ever stay in such a contradictory relationship with the categorical imperative since this imperative has no content, being formal. It does not demand you to behave in a particular way but *in accordance* with a general understanding of nature. What you must not contradict is the idea of the law of nature (somehow in the sense of not violating a natural law), rather than the categorical imperative itself. This is why, first, you must transform a certain intention into a possible law of nature and then see if the nature that such a law makes possible could subsist or not. Thus, as even Kant himself states, you can verify the morality of a behavior precisely through its compatibility or logical unity with the *generalization* of that behavior.

³ John Stuart Mill, *op. cit.*, p. 10.

Suppose that I want to know if lying is or is not a moral behavior. I will make a generalization and transform conceptually the act of lying into a law of nature and imagine that all human beings are liars. Through such a generalization, I get an image of human society that makes it evident that it cannot survive and develop if its members are all liars. Now, does this conclusion result from comparing two propositions, of which one negates something that the other states? Certainly not. You understand the conclusion only by following, to a certain extent, empirically the consequences of such an imagined law.

To a certain extent, the procedure is similar to a scientific experiment. As with a scientific experiment, you want to know what the consequence of introducing a new factor or force into an existing system that works very well right now is. The new force is meant to become a future constant presence of that system. You cannot know beforehand how the presence of that force influences the system. You must observe the consequences that occur factually. Imagine in this respect that you introduce a new microbe into a person's body. Initially, there is no logical contradictory relationship between it and the body. They are only two different realities, perhaps existing beside each other but not interacting together. However, combining them shows that they are incompatible with each other.

In other words, you cannot have a priori knowledge about how the presence of a factor (in this case, the concrete moral behavior) influences a system (in this case, human society conceived of as a system of nature).

Unlike this Kantian approach, the utilitarian view of Mill – in which he imagines a thought experiment according to which there could exist informed persons or competent judges concerning all kinds of pleasure, from the most vulgar to the noblest – is based on the idea of contradiction and not of drawing a consequence. This is because, in Mill's view, these informed persons cannot choose either for themselves or for all others anything other than something that promotes the highest pleasure as the highest or the noblest good. This good cannot contradict the initial option because, according to the initial concept of the "informed person", such a person should possess the knowledge of all possible pleasures pertaining to a human being, and, if someone chooses something evil, this is only because he is not knowledgeable about all those possible pleasures.

Here, indeed, we have a relationship of contradiction and not a computation or an experiment as in Kant's ethics. That is to say, a relationship in which the deeds someone accomplishes contradict logically another content. The contradiction would be the following: the informed person is someone who knows everything about human nature (this is why, in fact, he can assess differently various forms of pleasure); he knows both every possible action that harms it and every possible action that creates a possible pleasure in him, together with all its consequences. Related to such complete knowledge, any action of a human being will be only a particular case of that universal knowledge. Therefore, here, an evil deed stands undoubtedly in a relationship of contradiction with that knowledge because it is something that contradicts the known fact that that behavior is bad (in other words, when a person accomplishes that deed, he believes that the deed is good, although

according to the universal knowledge of the informed person, the deed is bad). Here, we have a contradictory relationship between the maxims of those deeds similar to the contradictory relationship between the following two propositions: “Fire burns you” (a piece of knowledge pertaining to the universal knowledge of the informed person) and “Fire does not burn you” (a claim asserted by an ignorant person who does not know the nature of fire and mocks the wisdom of those who say that fire burns). He resembles the morally ignorant person thinking that by doing something bad, he can reach something good, lacking thus the empirical knowledge of the competent judge who knows that bad deeds cannot lead to valuable pleasures). (We can notice in this respect that Mill’s utilitarianism agrees with Socrates’ view that doing an immoral act is clear proof of ignorance.)

In Kant’s ethics, you do not have a piece of prior knowledge; you must derive that knowledge through calculation, in the same way in which a scientist derives the result of the interactions of several forces by calculating the resultant force after he puts all those forces together.

2. HISTORY AND THE CONCEPT OF “COMPETENT JUDGES”

Another erroneous assumption concerning the concept of this “informed person” or “competent judge” is that such a person can be thought of as reaching his/her universal knowledge simply by experience, namely by letting things happen to him and then seeing what his emotional reactions are. To assume that rationality is a feature that develops by itself in the human being, similarly to how branches grow in a tree, is a fundamentally false conception of the human being. Rationality always implies an act of selection, too. In other words, we never let experiences come to us, but we choose those experiences based on what we have learned or have been taught. We proceed in this respect a priori by necessarily selecting experiences without having had any previous knowledge about those experiences. That means that all ‘noble’ pleasures also involve a negative moment, something like a feeling of shame concerning what is not noble. This denotes an internal constraint to be different than pure nature would want us to be.

The concept of the ‘informed person’ assumes thus something that cannot exist or occur. It assumes that experiences always *happen* to humans without them having any previous assessment of those experiences, and that only after having those experiences can humans draw the conclusion that that experience is good or beneficial – in which case it must be promoted – or bad, and, therefore, it must be repressed. For example, it assumes that humans, after having an inferior or vulgar pleasure, were attracted to the marvels of the celestial vault and discovered that its contemplation is a much more valuable experience than the indecent pleasure. Although this could be true at the individual level, it cannot happen at the historical level. Namely, those who discovered once that contemplation is a higher form of pleasure will tend to transmit this experience to their offspring. Then, these offspring taught by their parents that the celestial vault is a better experience than an inferior pleasure do not choose themselves but are taught to choose in that way; their parents

shape their sensibility, and they no longer have a pure, unbiased experience. They will not select freely but will tend to follow their parents' advice and assess the "inferior" pleasure as something less valuable.

The next step in the evolution could be that one sees a relationship between the lower pleasure and the celestial vault. For example, that there are some cosmic rhythms that can be felt simultaneously with the inferior pleasure. Suppose the parent or an older person teaches a younger person how to attain that feeling. In that case, both the young person and his offspring will never again assess the "natural" lower pleasure in the same way as something purely natural and spontaneous. Thus, humanity is always fundamentally biased in its experiences, and the parents' whims constantly distort the natural experiences of their children and, therefore, of all other descendants. The so-called "informed persons" would, therefore, never be unbiased but products of the history of their own communities.

We see thus that what Mill imagines is a class of people who enjoy complete freedom and thus are also ready to submit themselves to all kinds of experiences concerning pleasure. However, experiences are not free but are always biased. Mill seems to think that from whatever experiences humans have, the more pleasurable experiences will always come to the fore and ultimately be preferred by humans, as in a sort of backlash. Thus, it might be possible that during evolution, humanity constantly gets rid of the bad experiences and embraces the good experiences. However, human history shows something completely different, at least in one case, the victory of Christianity against the Roman Empire. Christianity introduced a certain world of values whose confirmation was based not on an immediate experience but on a promise. Christians did not stumble upon higher, nobler values but believed in them first. And this belief precedes, in Christianity, the knowledge of any informed person because it is related to something that cannot be known by any experience, namely the hope in a particular afterlife.

3. WHY "SUPERIOR" PLEASURES ARE NOT ONLY PLEASURES

The competent judges would be those who placed above history, might look down on history and see the consequences of each kind of pleasure humans tried out during history. The assumption somehow assumes that history is only a sort of succession of events that allows different types of pleasure to surface without interfering at all with their makeup. As if someone who commits a murder and afterwards happens to listen to a Mozart symphony, finds the music more pleasurable than the act of murdering. But such an assumption ignores the act of creating that music. And creating something that has an intellectual value or something that is preferred against lower pleasures involves a kind of focus that is not sustained by the energy of physical pleasure. It must be sustained by will. Animals are curious too, and they can both feel sexual arousal and curiosity. Still, they do not enter history, i.e., they are not able to transform their curiosity into something that is more than a momentary mood.

The capacity to transform curiosity into a constant feature of one's being – which is the equivalent of thinking, the intellect, or the spirit – does not result from its simple presence at the level of animality. Continuing to be curious (and I emphasize “continuing”), being able to maintain your state of astonishment which invites you to experience new sensations, is a feature that is irreducible to the pleasure of a natural being thought of as guided only by its impulses like an animal. No animal is continuously curious, or, expressed differently, no animal can *sustain* its curiosity through its will. This is exactly what happens with humans. They can maintain their curiosity – or, if you prefer, their focus – attached to an object as long as they want. The reason for this focus cannot be reduced to animal or natural interest, precisely because it involves more than interest; it involves a different ontological feature, namely will. And will is what Mill forgets in his interpretation of the human being as a searcher for happiness. Will is a feature that animals (or, more generally expressed, any simple natural being) lack, and, as Kant saw, it is the true basis of morality. An animal either fights – being moved by his instinct – or runs away, being again moved by its fear. It cannot voluntarily overcome its impulse. In other words, you do not have first a higher pleasure and then the action of pursuing it, as Mill believes, but first, you need to be capable of pursuing something voluntarily in order to reach higher pleasures.

Thus, the act of pursuing does not follow from the act of preferring higher pleasures, as Mill claims; it precedes them. This is why Mill's ultimately logical approach to the human being, and his utilitarianism – when he thinks that morality as a behavior that is grounded both on higher pleasures and an act of will necessarily entailed by them – are logically false. All criticisms oriented against Mill's endeavor of deducing what must happen from what there is (especially Moore's criticism) are, therefore, right. They all see how incomplete Mill's understanding of the human being is. The human will, however much related to human passions, is not human passion or impulse. It is a different component of the human soul, as Plato already saw in antiquity, when he metaphorically described the human soul as a carriage pulled by two horses – one, the animal impulses, but the other the will. Of course, only the latter was able to listen to the rational part of the soul, to the driver of the carriage.