Kant's Categorical Imperative

John Elliot V

In this essay I will present Immanuel Kant's Categorical Imperative and demonstrate how this reductive and flawed process can lead the adherent to ridiculous moral positions. I will conclude that Kant's Categorical Imperative is useless and that his project of deriving morality from reason is intrinsically impossible.

Immanuel Kant (1724-1804 AD) was a prolific German philosopher of the Enlightenment period. Among his work was a subset concerning morality and moral agency, a part of which we shall consider here.

Kant developed the Categorical Imperative, which in its first formulation was given by him as:

> Act only in accordance with that maxim through which you can at the same time will that it become a universal law.

What this means is that before you do anything be sure that you are satisfied that your behaviour would be suitable for all rational agents in all circumstances. If you cannot satisfy yourself of that, then according to Kant, your prospective action is immoral.

Kant believed that the Categorical Imperative required rational agents to always tell the truth. He says this is so because if people could lie there would be no utility in language, because nobody could be believed. Firstly, I might note, Kant claims to be concerned with intentions, not consequences, so on what basis does he require that there be utility in language? That is a consequential utilitarian concern. Secondly, in the world as we find it, people do lie, quite often; nevertheless there is utility in language. Perhaps less utility than if there were no (deliberate) lies, but still more than enough utility to be practical. Sometimes you believe what you are told and you're right, sometimes you believe what you are told and you're mistaken (this is probably harmless in many cases, perhaps even beneficial in others), and sometimes you don't believe what you've been told (and are either right or wrong about that).

Regarding speaking the truth: who has the truth? Does a Mormon have the truth or a Muslim? Perhaps the only honest thing one could say is "I don't know." And where is the utility in language then, when the only thing that can be uttered is that we're not sure?

As rational beings in this universe we have access to rational and empirical modes of discovering "truth" (or knowledge):

Rationality is limited because it cannot produce information, it can only reduce premises to conclusions via theory. The selection of premises to apply, and theories to apply them in, is an inherently irrational activity; the best we can do is look to experience to guide us in such selection and application.

Experience is limited because it is specific and confined to relatively small spaces and durations with respect to the scale of the universe; we simply cannot fathom most things because they are so far beyond our experience -- for example the number of atoms in the universe; the time the universe has existed for; and the dynamics of stars, black holes, or subatomic particles.

But that's all we've got in the toolbox. Kant wanted to throw out empiricism and confine us to reason, which is an impossible project, because you cannot apply (or select, or test) reasons without experience or judgement (which is inherently subjective).

It might be worth a small interlude at this point to comment further on the limits of rationality. If you follow rationality you are apt to discover gremlins within it, owing to its own nature. For example the halting problem as defined by Alan Turing wherein it is discovered that it is impossible to determine if an algorithm will halt, and return an answer, or get stuck and enter an infinite loop thereby running forever; or the incompleteness problem as discovered by Kurt Godel which stipulates that a formal system can be proved either complete or consistent, but not both. Other issues with rationality could be taken to include paradox of any kind.

Back on the earlier topic concerning honesty of moral agents, Kant goes on to say that to deceive another is to deny their moral status as an end in themselves. But we could take issue with that premise, that all moral agents are to be treated as ends in themselves. As mentioned earlier we could qualify moral agents as of either evil or beneficent intent and make exceptions accordingly. Why insist that cooperation with an evil agent is moral by definition? We could equally select an altered definition which denied the necessity to cooperate with evil agents, even in consideration of the difficulties that might arise in making such judgements concerning the evil or beneficent status of agents.

It is in this context that the French philosopher Benjamin Constant posed the following (paraphrased) thought experiment for Immanuel Kant, known as the Enquiring Murderer problem:

Suppose you are at home and your distressed friend arrives at the door claiming that a man is chasing him to kill him. You let your friend in and he runs upstairs. After a second knock at the door you find the would-be murderer, and he asks where your friend is. What is your moral obligation?

Kant held that it is your moral duty to tell the would-be murderer that your friend is upstairs. This is on the basis that the Categorical Imperative implies that you need to always tell the truth without respect to consequences. It should be enough to note that his system is disproved on the basis of reductio ad absurdum. The system requires absurd behaviour.

But there are other issues. Let's say you've swallowed Kant's nonsense and feel you have to tell the murderer the truth. You tell the murderer that your friend is upstairs. The murderer burst into the house and runs upstairs. Meanwhile your friend is escaping downstairs out the back door. Unbeknownst to you your friend had come downstairs and run out of the house. How would we characterise this situation?

Another thing to be suspicious of regarding Kant's philosophy is that reality is *complicated*. By way of comparison let's consider what's known as Enterprise Software. Enterprise Software is that class of software run by government agencies, businesses and other bureaucratic organisations to help automate the conduct of their affairs. Let's consider a few hypothetical and probably oversimplified examples:

The admissions process of a university: entry to the course will be granted if the applicant's UAI is 92.1 or greater, except if they're female in which case it is 70.2 as we want to get more woman into this course; the fee is $827, except if they're an international student in which case it's $948; discount 20% if paid by cash in advance, set to zero if scholarship; and so on...

The interest on a bank account: 5% for a transaction or cheque account, 13% for a savings account; bonus 2% if promo code 0235 was applied during February; less account keeping fee of $5 for account balances (before interest) below $200, 1% otherwise; and so no...

That's the sort of stuff that Enterprise Software manages, and the important thing to understand is that it is *complicated*. In practice hundreds of people are employed to generate and maintain this software and it grows so complex that no one person understands the whole thing. If computers couldn't process the details many many orders of magnitude faster than a person could the whole system would be impossible.

Central Processing Units executing billions of instructions per second, running on hundreds of servers, for many hours generate results. Architects deal with general problems and composition of subsystems. Programmers deal with specifics of smaller modules. Analysts aggregate and correlate reams of data with the aid of Business Intelligence algorithms. Managers coordinate the development effort. It's a big job.

And there are interesting phenomena in such systems, namely, "bugs". The word "bugs" is jargon for software errors. Program code that does the wrong thing. In practice they happen all the time and no large software is free of them. It's an inherent problem with codification of complex rules. Mistakes are inevitable.

It seems to me that any formal system with enough fidelity to encode realistic moral systems would suffer all the same problems. It would be too big to understand, too onerous to process, and laden with subtle mistakes and over-generalisations.

In conclusion I surmise that Kant's programme of reducing morality to reason is hopeless. Reason alone is not adequate for the task. His Categorical Imperative leads to ridiculous conclusions, and there is no basis on which we should accept it.

Bibliography

Kant, Immanuel; translated by James W. Ellington [1785] (1993). *Grounding for the Metaphysics of Morals 3rd ed.* Hackett.

Immanuel Kant (Stanford Encyclopedia of Philosophy). plato.stanford.edu. Retrieved 2014-18-24.

Lakoff, George. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, Basic Books 1999.

On formally undecidable propositions of Principia Mathematica and related systems I, Kurt Godel, 1931

On computable numbers, with an application to the Entscheidungsproblem, Alan Turing, 1937

Patterns of Enterprise Application Architecture, Martin Fowler, et. al. 2003