Truth as an instrumental good

Diomides Mavroyiannis*

May 10, 2021

Abstract

We study the role of truth as an instrumental good. We find that if two agents are communicating and truth is viewed by one agent as an instrumental good, then the other agent only has a reason to trust them if they know they want the same thing.

*dmavroyiannis8@gmail.com

1 Introduction

The incentive compatibility problem of consequentialism.

Utilitarianism and the hive mind.

Consider the trolley problem. Now let us modify this problem a little bit, now, we are no longer the ones making a decision. Greg is going to make a decision. Unfortunately, the lever has been replaced by two buttons, one button will lock the tracks so that it will go run over the one person, the other button will shift the train. Unfortunately, the buttons are not labeled. However, you know which of these buttons will save 5, do you lie to the person?

In this paper I wish to argue that utilitarians have no reason to tell the truth to non-utilitarians and as a result, there is no reason for a non-utilitarian to believe a utilitarian.

It is also argued that a utilitarian ALSO has no reason to tell the truth to anybody who has a different epistemic standard than themselves. For instance, it may be that both agree that utility may be maximized, but if the agent's disagree about the effects of actions, then a utilitarian wants to lie. There may be an exception to this if the utilitarian has epistemic humility of a very high degree but there is no reason.

Anyone agent who holds a moral code which views truth as merely an instrumental good will have a hard time being credible and being given any non-transparent role.

To facilitate discussion, let us use the sender - receiver distinction made in economics. One agent, the receiver, is about to take an action, and the other agent, the sender, is communicating to the receiver, information about the effects of those actions. We say that the agents are communicating when there is a genuine transfer of information from the sender to the receiver.

First consider the case where the two agents have the same consequential goals, the sender has no reason to lie. That is, I assume that communicating in itself, will not change the outcomes. This excludes a case where if the receiver acted on his own, he would perhaps learn a new skill, this would be a different consequence.

Now consider the case where the two agents have different ends. Suppose a scientist with a Singerian moral framework was investigating whether some animal has a different capacity

If the agent who is sending claims to hold consequential ethical frameworks, then they are morally obligated not to communicate with agents who may use that information to NOT increase total consequences in the way the sender envisions it.

Indeed, if for some reason the sender has credibility from the point of view of the receiver, then the sender is morally obligated to lie. If receiver is looking for her child, but the sender things consequences would be better if she did not find her child, then the sender is morally obligated to lie to the receiver.

However, the problem is exactly one of credibility. Why would the receiver believe the sender if it

is known what the ethical framework of the sender is?

Utilitarianism and other systems of ethics which propose measurements to be optimized all suffer from a well known problem in economics. Cheap talk is what happens when people are sending each other information that cannot be verified.

When the signal cannot be verified, the most important information the receiver can have is information on what the other person is optimizing.

If the receiver goals are aligned with the sender, then no issue arises. The sender has no reason to lie and has the incentive to be as truthful as possible.

If on the other hand, their goals are always opposed, the sender never has any reason to give any truthful information to the receiver.

If it is verifiable then we have a Bayesian Persuasion case

Throughout this paper we will assume that communicating is relatively costless. This does not have to be strictly true, it just has to be that people generally don't mind communicating, either because they enjoy it or because they find it virtuous or any reason.

The verifiability of statements is a property about future states of affairs. A statement is verifiable if some action can be taken that allows corroboration of the statement. Though this was initially thought to be the only class of statements that have information, by say Ayer or Popper, this view has since been abandoned. Nevertheless verifiable statements still remain, but what is especially interesting here is the non-verifiability of statements.

Here we are not only concerned with statements that are not verifiable in principle but also statement that are not practically verifiable. For instance 'I went to the store' may be a non-verifiable statement if nobody has seen you go down there.

When experts are communicating non-experts there is also an issue of costly verifiability. If for instance the expert is making statements to the layman that the layman would have to study years to confirm or verify then that information is in practice not verifiable by the layman. But this lack of verifiability, in fact changes the incentives of the expert.

Whether somebody can get away with lying depends on their interlocutors ability to verify. As such, if there is a more symmetric informational situation, then there is less information that cannot be verified. If there is assymetry, perhaps one is a professor at a high ranking institution, and the other is a business owner, then there seems to be almost no reason for the professor to be truthful.

Indeed, almost any consequentialist theory of ethics has a very limited reason to truthfully communicate.

Consequentialist theories of ethics, and here I have in mind mostly utilitarians and egalitarians, do have some reason to tell the truth. If the truth can be verified, then one has to worry about future shifts in opinions about truth. Lying about a truth that can be verified is often counterproductive because once it is verified, the liar will be exposed, and hence his future ability to lie will be

hindered. If on the other hand, the truth cannot be verified, the utilitarian has a license to lie at will.

But consequentialist theories have no reason to worry about non-verifiable truths. Indeed, if some truths can in fact increase the consequences they purport to prioritize then there is good reason to suppress or even create false evidence.

2 Example

What would an ethical lawyer do? Let's say that one of your clients has died. You have two wills they left behind. One will he wrote when he was 50, he said he is utilitarian and wants to give all his assets to the most effective charity. The other will, he wrote when he was 60, changed his mind and wants to give all his assets to his family. The family is not aware he wrote two wills, they only know he has at least one. It seems like a utilitarian lawyer would, in fact, NOT respect the wishes of the dead and simply show the family the first will.

If somebody has an ethical system that does not put truth as something that is good in itself, then there is no reason for them to communicate.

Truth as an instrumental good

3 The line between organisms

Consider a grandfather, his only pleasure in life is imagining the continuation of his grand kids. If there is an equilibrium where somebody has the power to re-direct his assets to others, then he would be unhappy. In other words, if a significant portion of peoples happiness is imagining their future offspring living a certain life, then a utilitarian must focus on commitment devices to stop utilitarians from redirecting resources in the future.

Indeed, perhaps the best commitment device, is stopping the spread of utilitarianism altogether.

4 Can you trust a utilitarian?

Is there any way that somebody can signal that they are in fact non-utilitarians? The only way is by giving a utilitarian significant power and they use that power to signal that they are not utilitarian.

For instance, it may be that they take numerous non-utilitarian actions, to signal to people that they are not utilitarians SO that they can eventually take a utilitarian action. The ONLY way to show that you are not a utilitarian is to take NON-utilitarian action on the first half of the actions you can take.

Since the personal bond usually means nothing relative to the global optimal, this means that there is no sense actually bonding with a utilitarian. Their propensity to help you is independent of your bond.

However, it may be worthwhile to get a utilitarian to have a good opinion of you. If for some reason the utilitarian cannot spend their time improving the global optimal and can spare some time for themselves, then it seems like it would make sense to try to get the utilitarian to like you. If they are too weak to influence other's happiness, then they can at the very least influence their own.

If somebody owns more than 50 per cent of the wealth, then that person can never have a sufficiently good signal.