Gray Zone Modal Utility Transport

by Sven Nilsen, 2020

In this paper I represent a method to transport utility of ethical problems in a compatible way with utilitarianism that has some properties of "gray zone" modality.

In the Trolley problem, a man can choose to pull a lever to save five lives at the cost of one life:

(-5, 1) vs (5, -1)	Utility vector of choices

In utilitarianism, a technique to simplify solving such dilemmas is to use equivalent utility transport. One can add or subtract lives for every choice, as long the same number of lives is added or subtracted. If the dilemma gets easier to solve in any of those other equivalent transformed problems, then the decision that holds for the easiest problem also holds for seemingly more difficult problems.

However, one would like to capture some properties of what is meant by "difficult problems".
The problem with equivalent utility transport is that it completely ignores this difficulty.
Only because one can determine the same optimal decision in two equivalent situations, does not mean that there is no way to compare the two situations and evaluate which one is counter-factual preferred.

This means that one would like to introduce some modality semantics into utilitarianism.
In modal logic, there is a "possible worlds" semantics, which will also be the case here.
To model this semantics, I make the following assumptions:

1. There exists a default choice
2. There exists a best possible world when the default choice is maximum utility
3. There exists a worst possible world when the default choice is minimum utility

The factor of best vs worst possible world is measured by a value `gz` which stands for "gray zone".

gz = 0	Best possible world
gz = 1	Worst possible world
0 < gz < 1	Gray zone possible world

Hence, the semantics using `gz` is a gray zone modal utility theory.

In order to create a utility transport over this semantics, one must add or subtract something for every choice, which will not change the optimal decision, therefore is consistent with utilitarianism.
However, instead of transporting to the simplest equivalent problem, the purpose of this new utility transport is to determine an equivalent problem that models the same "ethics difficulty".

utility – minimum	The "easiest" equivalent problem (old)
utility – gz · maximum	The "a-priori ethics difficulty" equivalent problem (new)

The gray zone modality utility transport is designed to counter-weight equivalent utility transport.

Here are some properties of the gray zone modality utility transport:

- It is consistent with utilitarianism
- Preserves homogeneity of degree one `∀ a { f(a · x) = a · f(x) }`
- Does not preserve additivity (therefore does not preserve linearity)
- In the best possible world, utilities are unchanged
- When maximum utility is zero, utilities are unchanged
- When switching to a lesser evil, negative utility is “mercy discounted”
- When switching to a lesser evil from a gray zone world,
  it is not possible to get zero (100% mercy discount)
- All dilemmas (two choices) are either best possible world or 100% mercy discounted
- Mercy discounting decreases under repetitive rational decision making

A “switching to lesser evil” situation occurs when:

- It is not the best possible world...
- … and the maximum utility is less than zero.

A mercy discount is when a negative utility is transported in positive direction.

In the Trolley problem, there are only two choices. The switch operator is in the worst possible world, since the default position of the lever gives minimum utility. When switching the lever, the operator is 100% mercy discounted the negative consequences from making that choice.

However, by adding a third track with 101 people to the Trolley problem:

[-5, -1] => [-4, 0]	Standard Trolley problem (100% mercy)
[-5, -1, -101] => [-4.96, -0.96, -100.96]	Small mercy discount (4%, gz = 0.04)

This happens because this gray zone world is closer to the best than the worst possible world.

The interpretation of comparing the situation of two choices versus three choices is ambiguous.

However, look what happens when one increases the number of people on the default track to 100:

[-100, -1, -101] => [-99.01, -0.01, -100.01]

It is much preferable to be in this gray zone world than the previous one.
Here, the switch operator almost get a full mercy discount.

If it is possible to imagine a world which things could have been very bad, and this world is close, then by making good decisions to maximize utility that brings the world close to the best possible one is considered an easy ethical decision. However, when you get closer to the best possible world, sacrificing a few for the many is considered more and more ethically difficult.

The gray zone modality utility theory permits imaginary choices to evaluate how ethically difficult a problem is, compared to the best or worst possible world. Therefore, it can measure the “gray zone”.