Preregistration

Ends versus Means: Kantians, Utilitarians and Moral Decisions

Aligned Robustness Experiment

Roland Bénabou   Armin Falk   Luca Henkel

November 21, 2023

1   Project Description

This experiment is an additional robustness experiment complementing the main experiment of the paper "Ends versus Means: Kantians, Utilitarians and Moral Decisions". In the main experiment, subjects face a series of ends-versus-means (EVM) dilemmas. Each dilemma features an ends-versus-means tension between the moral principles of consequentialism and deontological ethics. In the robustness experiment, we alter the dilemma situations so that they no longer feature a tension between the two principles while leaving the instructions otherwise identical. We call these decisions *aligned*, because, in contrast to the main experiment, the two moral principles now align in their prescription of the moral choice. Doing so allows us to investigate the following research question:

- Are our results of the main experiment an artifact of subjects' confusion or limited attention to the decision's consequences, or do they reflect deliberate and conscious decisions?

2   Experimental Design

The experiment consists of a single session in which subjects are confronted with seven distinct decision situations. The details of the decision situations are explained in section 3.

1. Trolley problem aligned (3.1)

2. Control donation decision (3.2)

3. Group donation game aligned (3.3)

4. Lying game aligned (3.4)

5. Bribing game aligned (3.5)

6. Statement choice aligned(3.6)

7. Rule following task aligned (3.7)

The order in which subjects face the decision-situations is randomized. For those decision situations that are (potentially) difficult to understand, subjects are asked comprehension questions.

2.1 Incentives

Subjects are paid a fixed payment for completing the entire experiment and can earn additional money through their decisions. If they do not successfully complete the session, they will not receive any payment. This feature is intended to minimize attrition. All decisions except the aligned trolley problem are incentivized: at the end of the experiment, one decision of each subject is randomly selected and implemented with real consequences.

2.2 Setting and Sample Size

The experiment will be conducted online with subjects from the subject pool of the BonnEconLab. Most of them are students from various fields of study. The targeted sample size is about 120 subjects. Since we expect some attrition, slightly more invitations will be sent out in order to achieve the intended sample size approximately.

2.3 Criteria for Excluding Subjects

We use the same exclusion criteria as in the main experiment: We will exclude any subjects that do not fully complete the experiment. Furthermore, we will exclude the top 1% of subjects in the response time distribution. Finally, we will exclude subjects that choose Option B in the control donation decision, see section 3.2 for the details of the decision.

3 Details on the Decisions and Hypotheses

The following sections will give details on the decision situations. For each situation, we briefly describe the situation and the options available to the subjects. We then state, based on a broad interpretation of the two concepts, what deontological ethics and consequentialism would prescribe as the morally right choice in the given situation.

3.1 Trolley Problem Aligned

We implement a version of the trolley problem using the Saving a Life Paradigm (Bénabou et al., 2020; Falk and Graeber, 2020). Following their considerations, we determined an amount of 380 €, which allows the Indian non-profit organization *Operation ASHA* to treat five people suffering from tuberculosis in India. In expectation, this saves one life from death by tuberculosis.

For the trolley problem, we have initiated a donation sufficient to save three human lives from tuberculosis in Indian state A (treat fifteen patients). There are also other people suffering from tuberculosis in another Indian state B, however. Subjects in the experiment can choose to redirect the donation from state A to state B. If redirected, the amount will be reduced so that the amount is sufficient to save one person in state B (treat five patients), again in expectation. Subjects can choose whether to redirect the donation using an animated graphic, where they have to move a slider in order to redirect the donation. Subjects will be incentivized to read the instructions explaining the trolley situation carefully: they are notified that they will be asked four comprehension questions after their decision. They will receive 0.5 € at the end of the experiment for each correctly answered question.

Moral choice. Deontological ethics: Do not redirect. Consequentialism: Do not redirect. Hence, the aligned choice is *Do not redirect.*

3.2 Control Donation Decision

This decision situation serves as a control for the next situations, establishing subject's basic preferences between donations to a good cause and increases in payments to other subjects. Subjects choose one of two Options A and B. Option A triggers a 15 € donation to the *Förderkreis für krebskranke Kinder und Jugendliche*, which is a charity located in Germany devoted to the support of children who suffer from cancer. Option B will anonymously increase the payment of one other subject participating in the experiment by 2 €.

Moral choice. Deontological ethics: Option A. Consequentialism: Option A. Hence, the aligned choice is *Option A.*

3.3 Group Donation Game Aligned

In this decision situation, subjects are grouped with five other subjects in a game adapted from Falk, Neuber, and Szech (2020). The other group members are subjects recruited from the subject pool of the BonnEconLab who do not take part in the main experiment. The group is entrusted with a 2 € bonus payment to another subject (not part of the group). Subjects choose sequentially between two options, A and B. Choosing Option A triggers a 15 € donation to the *Förderkreis für krebskranke Kinder und Jugendliche*. Choosing Option B has no further consequences. If at least one of the six group members chooses Option A, the bonus payment is destroyed, i.e., not carried out. Subjects of the main experiment choose last and are informed about whether the bonus payment has already been destroyed. If so, their choice is non-consequential with respect to the bonus payment. Based on previous experience, we expect that in each group, at least one of the first five subjects chooses Option A. Hence (almost) all subjects will face the situation in which the bonus payment to the other subject is destroyed. For them, the following hypothesis applies:

Moral choice. Deontological ethics: Option A. Consequentialism: Option A. Hence, the aligned choice is *Option A.*

3.4 Lying Game Aligned

In this decision situation, subjects play a sender-receiver game adapted from Gneezy (2005) in the role of the sender. Receivers are other subjects recruited from the subject pool of the BonnEconLab who do not take part in the main experiment. Receivers face two Options, A and B, but do not know the corresponding consequences. Senders know that Option A will trigger a 15 € donation to the *Förderkreis*, and Option B will increase the payment of the receiver by 2 €. Senders can send one of two messages to the receivers. Message 1 contains the statement that Option A will trigger a higher donation (the truth), while Message 2 states that Option B will trigger the higher donation (a lie). They are truthfully informed that in such a situation, the receivers will, with an overwhelming majority, choose the option that the message suggests will trigger the higher donation.

Moral choice. Deontological ethics: Message 1. Consequentialism: Message 1. Hence, the aligned choice is *Message 1.*

3.5 Bribing Game Aligned

In this decision situation, subjects (called senders) can choose between the two options A and B, as in section 3.2. However, in this situation, the choice has additional consequences. Prior to making their choice, they are informed that the receiver of the 2 € of Option B will afterward decide between two other Options C and D. Choosing Option C will trigger a 20 € donation to the *Förderkreis* while Option D will yield the receiver 2 €. Importantly, the receiver has determined (binding) to choose Option C only if the sender selects the donation by choosing Option A and will choose Option D otherwise. The receivers take part in a separate experimental session conducted before the main experiment and are matched to the subjects of the main experiment when they decide to condition their choice in the previously described way. Thus, senders face the choice of whether to choose Option A, which will result in the receiver choosing Option C or Option B, after which the receiver will choose Option D.

Moral choice. Deontological ethics: Option A. Consequentialism: Option A. Hence, the aligned choice is *Option A.*

3.6 Statement Choice Aligned

In this decision situation, there exists a CO_2 certificate, which, independently of subjects' decisions, is destroyed with a probability of 50%. Subjects can press one of two buttons, each linked to a different statement related to the environment. Button 1 states "I support the preservation and protection of the environment", and Button 2 states "I support the destruction of the environment". Choosing to click on Button 2 has no further consequence. Choosing Button 1 instead will result in a 15 € donation to the *Förderkreis* in case the certificate gets destroyed. Before this decision, subjects are asked about their attitudes towards the environment using two Likert scales.

Moral choice. Deontological ethics: Button 1. Consequentialism: Button 1. Hence, the aligned choice is *Button 1.*

3.7 Rule Following Task Aligned

This decision situation builts on the Rule-Following task from Kimbrough and Vostroknutov (2016) to measure individual normsensitivity with one important difference: subjects are paid based on the time the stick figure walks across the computer screen, with more time yielding a higher bonus payment. That is, if subjects follow the rule (wait until the lights turn green), they are rewarded with a higher payment.

Moral choice. Deontological ethics: Do not break rule. Consequentialism: Do not break rule. Hence, the aligned choice is *not breaking the rule.*

3.8 Hypothesis

Based on our results from the main experiment, our hypothesis for the robustness experiment is that the large majority of subjects choose the aligned choice in each decision situation.

References

Bénabou, Roland, Armin Falk, Luca Henkel, and Jean Tirole (2020). "Eliciting Moral Preferences: Theory and Experiment". *Working Paper.*

Falk, Armin and Thomas Graeber (2020). "Delayed Negative Effects of Prosocial Spending on Happiness". *Proceedings of the National Academy of Sciences* 117.12, pp. 6463–6468.

Falk, Armin, Thomas Neuber, and Nora Szech (2020). "Diffusion of being pivotal and immoral outcomes". *Review of Economic Studies* 87.5, pp. 2205–2229.

Gneezy, Uri (2005). "Deception: The role of consequences". *American Economic Review* 95.1, pp. 384–394.

Kimbrough, Erik O. and Alexander Vostroknutov (2016). "Norms Make Preferences Social". *Journal of the European Economic Association* 14.3, pp. 608–638.