<div align="center">

Preregistration

Perceptions and Realities: Kantians, Utilitarians and Actual Moral Decisions

Roland Bénabou   Armin Falk   Luca Henkel

June 28, 2021

</div>

1   Project Description

This project concerns moral behavior and how observed choices relate to moral orientation. We study experimentally situations where the two most influential moral theories, consequentialism and deontological ethics, distinguish themselves sharply on their directives regarding the morally right action. In particular, one such situation is the so-called "trolley problem" (Foot, 1967; Thompson, 1976), which we implement both with real and hypothetical consequences. By observing behavior in the different situations, we investigate how choices vary between the different situations, as well as how behavior in these situations relates to general economic preferences and judgment obtained from hypothetical questionnaires. Doing so will be informative about the manifestation of consequentialism and deontological ethics as individual moral orientation in choice behavior. Hence, we ask the following broad research questions:

- Are there moral types that behave consistent with consequentialism and deontological ethics and can they be identified by observing choices in the trolley problem or other situations? Does this depend on whether decisions come with real consequences or are posed hypothetically?

- Is consequentialistic and deontologically motivated behavior correlated with general economic preferences and other relevant dimensions?

2   Experimental Design

The experiment consists of two sessions that are separated in time by one week. All subjects have to complete both sessions. In each session, subjects are confronted with a set of distinct decision situations complemented by questionnaires. The following list gives an overview of the content of the experiment, with the details explained in sections 3 and 4.

- Main decision situations

    1. Trolley problem (3.1)

    2. Control donation decision (3.2)

    3. Group donation game (3.3)

    4. Lying game (3.4)

    5. Bribing game (3.5)

    6. Statement choice (3.6)

- Additional decision situations

    7. Rule following task (3.7)

    8. Intention vs. Outcome game (3.8)

    9. Trust game (3.9)

    10. Public goods game (3.10)

    11. Dictator game (3.11)

- Additional items

    12. Attention checks (4.1)

    13. Questionnaires (4.2)

The trolley problem (3.1) and the other main decision situations (3.2 - 3.6) will appear in separate sessions. The order is randomized, i.e., one-half of subjects start with the trolley problem in the first session and then face the other main decision situations in the second, while for the other half, it is the other way around. Within each session, decisions are randomized in order as well. The questionnaires are distributed between sessions to make both sessions approximately of equal length. For those decision situations that are (potentially) difficult to understand, subjects are asked comprehension questions.

2.1 Treatments

There is one between-subject treatment variation, which concerns the trolley problem, resulting in two treatments. In treatment *Hypothetical*, the trolley problem is posed hypothetically, that is, without actual consequences. In contrast, in treatment *Real* the decision is potentially implemented: for each subject, there is a 10% probability that the subject's decision is carried out exactly as stated. Whether their choices have a 10% chance of having real consequences or not is announced to the subjects before they receive detailed information about the decision. Apart from this announcement, the treatments are identical in content. Thus all other content and analyses will be within-subject.

2.2 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 both sessions, they will not receive any payment. This feature is intended to minimize attrition. For each session, one decision (not including the trolley problem decision) is randomly selected and implemented with real consequences. Subjects only learn at the end of the experiment which decisions were selected and what their payoff was. Additionally, in the *Real* treatment, the decision in the trolley problem is potentially implemented with real consequences, as described before.

2.3 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 600 in total, 300

for each treatment. Since we expect some attrition, slightly more invitations will be sent out in order to achieve the intended sample size approximately.

2.4 Criteria for Excluding Subjects

The following exclusion criteria will apply (in the presented order): We will exclude any subjects that do not fully complete the experiment. In order to complete the experiment, subjects have to successfully complete both sessions. Furthermore, we will exclude the top 1% of subjects in the response time distribution over both sessions. 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, on the one hand, and consequentialism, on the other hand, would prescribe as the morally right choice in the given situation.

3.1 Trolley Problem

We implement a version of the trolley problem that can be conducted with real consequences. In order to do so, we adapt the previously used 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 one human life from tuberculosis in Indian state A (treat five 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 multiplied by three so that that the amount (now 1,140 €) is sufficient to save three people in state B (treat fifteen patients), again in expectation. This closely mirrors the decision made in the trolley, but in a realistic and feasible way, with real lives saved in the process. 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.

We will use the between-subject treatment variation to assess how real consequences vs. hypothetical decision-making influences choices in the trolley problem. We will investigate how behavior in the trolley is related to choice behavior in the other situations that are presented next, as well as to questionnaires that focus on moral judgment.

Moral choice. Deontological ethics: Do not redirect. Consequentialism: Redirect.

3.2 Control Donation Decision

This decision situation serves as a control for the next three 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. The decision is designed to have with Option A a clearly morally preferable option under both deontological ethics and consequentialism. We expect the vast majority of subjects to choose Option A. Since the next three decisions rely on subjects preferring Option A in this decision, we exclude those subjects from the main analysis that choose Option B, as described in section 2.4.

Additionally, in order to assess subject's perceived moral relevance of the choice, we will ask subjects whether they think that there exists a morally right and wrong option (1) for them and (2) for other people. Here we again hypothesize that subjects perceive Option A clearly as the morally right option.

3.3 Group Donation Game

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 15 € donation to the *Förderkreis* and chooses sequentially between two options, A and B. Choosing Option A has no further consequences. For subjects of the main experiment, choosing Option B will generate 2 € for another subject. When other group members choose Option B, their own payoff increases by 2 €. However, if at least one of the six group members chooses Option B, the donation is destroyed, i.e., not carried out. Subjects of the main experiment choose last and are informed about whether the donation has already been destroyed. If so, their choice is non-consequential with respect to the donation. Based on previous experience, we expect that in each group, at least one of the first five subjects chooses Option B. Hence (almost) all subjects will face non-consequential donation situations. For them, the following hypothesis applies:

Moral choice. Deontological ethics: Option A. Consequentialism: Option B.

3.4 Lying Game

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 B will give a higher personal payment (the truth), while Message 2 states that Option A will give the higher payment (a lie). They are truthfully informed that in such a situation, the receivers will, with an overwhelming majority (> 90%, obtained from a short pilot), choose the option that the message suggests will give them more money. Senders

thus face the choice of whether to lie to the receivers in order for them to choose the option that will trigger the donation.

Moral choice. Deontological ethics: Message 1. Consequentialism: Message 2.

3.5 Bribing Game

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 yield the receiver 2 €, while Option D will trigger a 20 € donation to the *Förderkreis*. Importantly, the receiver has determined (binding) to choose Option D only if the sender sends 2 € to him/her by choosing Option B, and will choose Option C 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 B.

3.6 Statement Choice

In this decision situation, there exists a CO_2 certificate, which, independently of subject's 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 1 has no further consequence. Choosing Button 2 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 2.

3.7 Rule Following Task

This decision situation uses the Rule-Following task from Kimbrough and Vostroknutov (2016) to measure individual normsensitivity. The task will be implemented exactly as described in their paper. The game will be used to assess the relationship between a general tendency to follow rules or norms as elicited by the game and the extend of behavior consistent with Deontological ethics in the previously described decision situations.

3.8 Intention vs. Outcome Game

In this decision situation, subjects make four dictator game decisions as dictators. In each, they can distribute 10 € between themselves and another subject, the receiver. Each amount distributed to the receiver is tripled. Before dictators make their distributional choices, the receiver chooses between two lotteries: Lottery M pays with a probability of 70% 10 € and triggers with 30% probability a donation

of 15 € to the *Förderkreis*. Lottery D has the probabilities reversed, triggering the donation with a probability of 70% and paying 10 € with a probability of 30%. Since receivers can choose between two options and there exist two potential outcomes, there are four potential possibilities with respect to chosen option and realized outcome. Using the strategy method, subjects make their distributional choice for each of the four possibilities. By allowing subjects to distribute differently conditional on the choice of the receiver and realized outcome, we can measure how subjects value/reward intentions compared to outcomes. The game will be used to assess the relationship between the propensity to reward intentions, and the extend of behavior consistent with Deontological ethics in the previously described decision situations.

3.9 Trust Game

Subjects play both roles of the standard trust game with an endowment of 5 € and a multiplication factor of 3. Senders can send integers from 0 to 5 €. Recipient choices are elicited using the strategy method, i.e., subjects decide how much to send back conditionally on each of the six possible amounts send. From the decisions in both roles, we get a measure of trust and positive reciprocity.

3.10 Public Goods Game

Subjects play the standard public goods game with an endowment of 5 €, a group size of 3, and a multiplication factor of 1.5.

3.11 Dictator Game

Subjects play two variations of a dictator game, a giving and a taking game. In each, they can allocate 20 € between themselves and the *Förderkreis*. In the taking game, 20 € are to be donated to the charity, and subjects can decide to take money for themselves. In the giving game, they are endowed with 20 € and can donate an amount to the charity. Using the giving game, we obtain a measure of prosociality. The difference between the two games gives a measure of the importance of the action (giving vs. taking).

4 Additional Items

4.1 Attention Checks

There are two attention checks within the questionnaire at the end of a session that serve as robustness checks for the situations described in sections 3.3 and 3.6. For the former situation, they are asked to recall how the other subjects in their group had chosen. For the latter, they are asked to recall whether there were additional consequences attached to the buttons. Both questions will be used to define each a binary measure (answer correct/not correct), which captures the extent to which subjects were attentive to the consequences in the two decision situations. These can then be used to assess the role of attention in driving observed behavior in the two situations.

4.2 Questionnaires

The following questionnaires are included:

- Moral Foundations Questionnaire (Graham et al., 2011)

- Oxford Utilitarian Scale (Kahane et al., 2018)

- Moral Universalism Module (Enke, Rodríguez-Padilla, and Zimmermann, 2021)

- Classical train track trolley problem

- Political attitudes and religiosity questionnaire

- Socio-demographic questionnaire: age, gender, subject of studies, final high school grade, gross monthly income, number of previous participations in experiments of the BonnEconLab, self-assessed understanding of instructions

References

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

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann (2021). "Moral Universalism: Measurement and Economic Relevance". *Management Science* forthcoming.

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.

Foot, Philippa (1967). "The Problem of Abortion and the Doctrine of the Double Effect". *Oxford Review* 5, pp. 5–15.

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

Graham, Jesse, Brian A. Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H. Ditto (2011). "Mapping the Moral Domain". *Journal of Personality and Social Psychology* 101.2, pp. 366–385.

Kahane, Guy, Jim A.C. Everett, Brian D. Earp, Lucius Caviola, Nadira S. Faber, Molly J. Crockett, and Julian Savulescu (2018). "Beyond sacrificial harm: A two-dimensional model of utilitarian psychology". *Psychological Review* 125.2, pp. 131–164.

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

Thompson, Judith Jarvis (1976). "Killing, Letting Die, and the Trolley Problem". *The Monist* 59.2, pp. 204–217.