Ends versus Means: Kantians, Utilitarians, and Moral Decisions

Roland Bénabou Armin Falk Luca Henkel

January 9, 2026

Abstract

Choosing what is morally right can be based on the consequences (ends) resulting from the decision – the Consequentialist view – or on the conformity of the means involved with some overarching notion of duty – the Deontological view. Using a series of experiments, we investigate the overall prevalence and the consistency of consequentialist and deontological decision-making, when these two moral principles come into conflict. Our design includes a real-stakes version of the classical trolley dilemma, four novel games that induce ends-versus-means tradeoffs, and a rule-following task. These six main games are supplemented with six classical self-versus-other choice tasks, allowing us to relate consequentialist/deontological behavior to standard measures of prosociality. Across the six main games, we find a sizeable prevalence (20 to 44%) of deontological choices by subjects, but no evidence of stable individual preference types across situations. Instead, which moral principle prevails appears to be context-dependent. In contrast, we find a substantial level of consistency across self-versus-other decisions, but individuals' degree of prosociality is unrelated to how they choose in ends-versus-means tradeoffs. In both types of decisions, subjects experience substantive normative uncertainty, as revealed by their willingness to randomize their choices when given the option to do so.

Keywords: morality, deontology, consequentialism, Kantian, ends versus means, trolley dilemma, altruism, social preferences, normative uncertainty, ambiguity aversion.
JEL classification: C91, D01, D64

Contact: Roland Bénabou: Princeton University, NBER, CEPR, IZA, and BREAD; rbenabou@princeton.edu. Armin Falk: University of Bonn; afalk@uni-bonn.de. Luca Henkel: Erasmus University Rotterdam, ERIM, CESifo, JILAEE, IZA, Tinbergen Institute; henkel@ese.eur.nl.
Acknowledgements: We are grateful for valuable comments to Jean-Francois Bonnefon, Valerio Capraro, Benjamin Enke, Brit Grosskopf, Pietro Ortoleva, Claudia Senik, Jean Tirole, and participants in seminars and conferences at briq, PSE, SciencesPo, TSE, UCEMA, Cornell, LSE, Imperial, and the University of Chicago. Egshiglen Batbayar, Erik Ortiz Covarrubias, Luca Michels, and Adam Rangel provided excellent research assistance. Bénabou gratefully acknowledges the hospitality of Sciences Po and the Paris School of Economics during the year 2021-2022.
Funding: Funded by the German Research Foundation (DFG) under Germany's Excellence Strategy - EXC 2126/1- 390838866. Support by the German Research Foundation (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged.
Ethics approval: The study obtained ethics approval from Princeton University (#12513 and #15939).
Research transparency: The study was preregistered at the AEA RCT Registry (#AEARCTR-0007714). See Appendix I on the mapping between preregistration and paper. Experimental instructions are in Appendix J.

1 Introduction

When providing a public good or engaging in reciprocal behavior, individuals trade off the costs and benefits to themselves against those to others. The question of what constitutes moral behavior then generally has a simple answer, given by the degree of prosociality of the chosen action. The preferences and image motives underlying such self-versus-other tradeoffs are by now fairly well understood. Much less studied is another important class of decisions, in which the question of what is the right thing to do elicits far less agreement. These are situations in which achieving some socially desirable *end* requires the use of *means* considered inherently objectionable. These include "sacrificial dilemmas" where helping or saving a greater number of people requires placing some in harm's way (e.g., hostage situations, medical triage, or vaccine priority), but also many less dramatic choices such as industrial and product safety, or bargaining situations where a bribe must be paid, a threat made, a lie told, or deeply held values compromised.

Unlike in "self-versus-other" choices, the two main approaches dominating Western moral philosophy and psychology now come into conflict. Under the Consequentialist view, the propriety of an action should be judged solely by its consequences (Bentham, 1789; Mill, 1863; Sinnott-Armstrong, 2019). Economics, in particular, is predominantly utilitarian, even when consequences include emotions or social image. The consequences of an act are evaluated and traded off within a decision-maker's preferences, and the social-welfare functions used for normative analyses are also consequentialist.

According to Deontological ethics, on the other hand, whether an action is permissible or not should be based only on its conformity with a series of rules, independent of their consequences for the situation at hand (Kant, 1785; Alexander and Moore, 2016). Many tradeoffs are then prescribed as violating higher values such as life or human dignity. While rarely absolute in practice, such principles do place major restrictions on both individual choices (lying aversion, sense of duty, refusals to "sell out") and the organization of society (taboos, proscription of "repugnant markets"). Relatedly, in the public debate, adherents of deontological views often criticize economists' use of cost-benefit analysis (e.g., value of a statistical life) as fundamentally out of touch with most people's moral intuitions.

Research questions. This paper aims to extend the empirical study of moral behavior and preferences beyond standard prosociality (altruism, reciprocity, fairness), into the other dimension of consequentialism versus deontological ethics. We do this by exploring three main questions.

First, how prevalent are these two moral principles? In particular, how many people exhibit consequentialist versus non-consequentialist tendencies in their actual behavior? If a significant fraction of people knowingly choose non-consequentialist options, representing their preferences in the standard manner will lead to inaccurate positive predictions and standard social welfare functions may be misleading guides to normative decisions.

Second, how consistent are these two moral principles? That is, can we characterize them as preferences that remain relatively stable across choice tasks – similar to risk-aversion, impatience, or especially altruism? Moving from between-subject variation to within-subject consistency, we examine how strongly an individual's behavior is correlated across *a range of situations* in which consequentialist and deontological reasoning prescribe opposing actions.[1]

Third, how is this other dimension of morality related to standard social preferences? Specifically, are deontologists generally more (or less) prosocial than consequentialists? Accordingly, we examine to what extent decisions pitting the two moral principles against each other may be correlated with those pitting self-interest against concern for others' welfare.

We start by presenting a simple model that formalizes these questions and frames our experimental investigation. We consider three nested sets of preferences. First, pure consequentialists always prefer actions that maximize overall consequences, while pure deontologists always prefer those that conform to a normative prescription and experience a psychological loss from violating it. Second, agents with mixed preferences put weight on both concerns, so their decisions will vary across choice problems. Third, we allow agents to be unsure of what their deep preferences are, or should be, and averse to such moral ambiguity. Indeed, there is growing evidence that people perceive substantial cognitive uncertainty even over basic aspects of their preferences, such as risk aversion and time discounting (Enke and Graeber, 2023; Enke, Graeber, and Oprea, 2023), and a fortiori over complex ones such as those pertaining to morality, i.e., they experience *normative uncertainty* (e.g., Dietrich and Jabarian, 2022; Ong and Qiu, 2023). Behaviorally, the model shows how the prevalence of consequentialist and deontological concerns is revealed through the distribution of choices in ends-versus-means dilemmas, their consistency can be assessed by analyzing the correlation of these decisions across different choice tasks, and the presence of normative uncertainty by examining the extent to which agents hedge by randomizing their choices.

Turning to the experimental core of the paper, our design consists of two main blocks of choice tasks: first, a novel set of tasks designed to identify consequentialist vs. deontological decision-making (and any corresponding stable preferences if they exist) through *ends-versus-means* (EVM) tradeoffs, and a second set capturing prosocial behavior through *self-versus-other* (SVO) tradeoffs.

Ends-versus-means choice tasks. Our first block starts with the trolley dilemma (Foot, 1967; Thompson, 1976), which asks whether it is permissible to sacrifice one life to save a larger number and has dominated normative debates and empirical investigations of moral decision-making. There are now large international surveys exploring different variants of it, aiming to provide guidance for decisions to be made by autonomous vehicles, robots, and other AI algorithms (e.g., Awad et al., 2018). Despite its popularity, the trolley dilemma has almost exclusively been posed as a *hypothetical* question, and cast in very abstract situations (another version being the organ-transplant dilemma). Our paper's first contribution is thus to experimentally study the trolley dilemma's value as a guide to the real decisions and policy choices it is meant to represent, by implementing a version of it with actual, life-saving consequences. Does this make a difference for behavior?

Our principal interest, however, is in the prevalence and the consistency of consequentialist versus deontological behavior across a broader set of less momentous, more common choice settings. The paper's second and main contribution is therefore to design a series of experimental games creating ends-versus-means tensions between these two moral principles, while leaving the decision maker's material and social payoffs unaffected. We start from a baseline situation in which the two principles

[1]This is a very different question from that of the temporal stability of moral judgments across successive repetitions of the same question, as we discuss in the conclusion.

align: if the choice is between directing a donation of 15 Euros to a recognized charity treating children suffering from cancer and one of 2 Euros to some random other subject, subjects almost universally agree that the former is the right thing to do. We then modify the decision problem so that bringing about this preferred social consequence requires breaking a normative rule, either: (i) *lying* to another subject, at their expense; (ii) *bribing* another subject, who controls greater resources; (iii) making a morally *repugnant* but entirely cheap-talk, anonymous, statement; (iv) refraining from "expressively" choosing a moral option when the choices of others have already made it *ineffective* and sticking to principle will only cause a further social loss; (v) violating an explicitly stated *rule,* absent any enforcement or meaningful effects of disobedience.

Self-versus-other choice tasks. Complementing this main series of six choice tasks, our second block consists of a rich set of familiar games measuring prosocial behavior, administered to the same subjects a week after the first block. These include a dictator game, involving either giving to or taking from a charity; a trust game, as both first- and as second-mover, using the strategy method; a public-goods game; and a moral-luck game (rewarding either intentions or outcomes). Finally, at the end of this block, we administer two questionnaires widely used in the moral-psychology literature, the Oxford Utilitarian Scale (Kahane et al., 2018) and the Moral Foundations Questionnaire (Graham et al., 2011), as well as the more recent Moral Universalism short module of Enke, Rodríguez-Padilla, and Zimmermann (2022); we also gather standard sociodemographic variables.

A key aspect of our design is that we measure choice behavior *within subject*. That is, our subjects face all choice tasks in randomized order, which allows us to look for stable patterns (preference types) within each of the consequentialist/deontological and prosocial/selfish dimensions, as well as for potential correlations between the two.[2]

We implemented three versions of the ends-versus-means choice tasks, conducted as separate experiments between subjects. In the first, main experiment ($n = 593$), subjects face the ends-versus-means tradeoffs, i.e., ends are in conflict with the means. In a second, robustness experiment ($n = 122$), subjects face a variant of the same choice tasks in which we simply reverse the payoff structure, so that the desirable ends now align with the permissible means. Comparing this *ends-mean-aligned* (EMA) condition with the main EVM one allows us to control for influences such as confusion or inattention in assessing the prevalence of consequentialist/deontological decision-making. In a third experiment ($n = 443$), we provide subjects with the explicit option to randomize between the option favoring the ends and the one favoring the means, following Agranov and Ortoleva (2023). This expanded choice set allows us to assess the prevalence of preference uncertainty and ambiguity aversion, as well as an additional test of consistency by studying the correlation of chosen probabilities across tasks. Our preregistered experiments were administered online as virtual lab experiments, with subjects being primarily students.[3]

Results. We report five main results. First, in every ends-versus-means choice task, a significant fraction of subjects choose non-consequentialist options that are in line with deontological principles.

[2]We can also use consistency of behavior in the well-known self-versus-other choice tasks to compare and benchmark consistency in the so far unexplored ends-versus-means choice tasks.

[3]https://www.socialscienceregistry.org/trials/7714. For details on the mapping between paper and preregistration, see Appendix I.

The fractions range from 20% in the bribing game to 44% in the repugnant-statement game. These revealed preferences clearly differ from those of "standard" economic agents. We employ extensive robustness checks to verify that the results are not just driven by limited attention or subjects' misunderstanding of the decision tasks. Most importantly, in our robustness experiment, where ends and means align, almost every subject chooses the aligned version, showing that they understand and pay close attention to the structure and consequences of the tasks.

Second, we deliver mixed news on the informativeness of the trolley dilemma as a guide to actual decisions. On the one hand, implementing it with real (statistical) lives at stake does not change behavior, relative to the standard procedure involving only hypothetical lives. In both cases, about 25% of choices are deontological. On the other hand, trolley choices predict close to no other behavior: across the eleven (EVM and SVO) other incentivized choice tasks, there is only a weak correlation of 0.16 with behavior in the lying game, and no significant correlation with any of the other ten, for which all correlations are below 0.06.

Third, and more generally, we find very little evidence of individual-level consistency in deontological versus consequentialist decision-making: choices across the six choice tasks pitting one principle against the other are largely uncorrelated. Almost no subjects choose the deontological option in five or more of the six choice tasks, only 10% in four or more, and only 10% never do so, behaving as standard economic agents; about 27% make one such choice, and 50% either two or three. Thus, there appears to be no single preference type that would robustly predict choices across ends-versus-means dilemmas. Instead, which moral principle prevails for an individual appears to be context-dependent. This is in stark contrast to standard social preferences, which exhibit substantial within-subject consistency across the six self-versus-other choice tasks in the second block. For instance, dictator-game giving reliably predicts behavior in all five other SVO choice tasks, with correlations ranging between 0.20 and 0.65.

Fourth, we report evidence consistent with substantial normative uncertainty. When given the choice to randomize, a significant fraction of subjects in every EVM choice task indeed do so. In fact, we show that the existence of an EVM tradeoff explains a significant share of randomization behavior: in the variants of the choice tasks where ends and means align (EMA condition), the share of subjects who randomize is substantially lower. However, while we find consistent randomization on the extensive margin -the same subjects tend to randomize in all tasks, as in Agranov, Healy, and Nielsen (2023)- there is no correlation across choice tasks in the randomization probabilities that subjects choose. This finding provides another piece of evidence for the lack of stable preference types in the deontological-consequentialist dimension.

Fifth, ends-versus-means decision-making is unrelated to general prosocial preferences. Put differently, acting deontologically versus consequentially in the trolley, or any of the other main EVM games, is not predictive of whether a person is more or less prosocial. This result contrasts with people's perceptions, which tend to rate individuals who make deontological actions as more moral and trustworthy (e.g., Everett, Pizarro, and Crockett, 2016; Capraro et al., 2018; Everett et al., 2018).

Finally, beyond its specific findings, our paper develops a flexible paradigm to study consequentialist vs. deontological decision-making, based on a new set of ends-versus-means choice tasks. These are fundamentally distinct from classical self-versus-other tradeoffs, and indeed reveal an essentially independent dimension of moral preferences.

Related literature. We first contribute to the nascent literature in economics that investigates behavior motivated by deontological versus consequentialist considerations. Starting with Laffont (1975), several papers have conceptualized Kant's Categorical Imperative for decision-making in strategic situations (Kranz, 2010; Roemer, 2010; Alger and Weibull, 2013, 2016; Alger, Weibull, and Lehmann, 2020). In these models, Kantian agents act (in part) as if their decision would cause all other players to make a similar choice, in line with the imperative's universalizing principle. Van Leeuwen and Alger (2021) find support for the existence of such a motive, mixed with standard Nash-consequentialist behavior, in the Prisoner's, Trust, and Ultimatum games.

Closely related is the literature on ethical voters, who derive a fixed "duty" or "expressive" utility from casting a ballot for an alternative that is considered morally superior, even if it goes against their self-interest or if the probability of being pivotal would not justify the cost of voting. Feddersen and Sandroni (2006), Feddersen, Gailmard, and Sandroni (2009), and Feess, Kerzenmacher, and Timofeyev (2022) develop such models; the last two also conduct experiments showing, by varying the probability that a vote will be pivotal, that both consequentialist and deontological or expressive motives are at work. Also related is the literature on rule-following, which shows that individuals conform to rules even without extrinsic incentives (Kimbrough and Vostroknutov, 2016; Gächter, Molleman, and Nosenzo, 2025). Finally, our paper is related to the literature studying the role of intention vs. consequences, showing that when judging the prosociality of an opponent's action, most individuals act non-consequentially by considering intentions: They reward or sanction actions yielding identical consequences differently if underlying intentions differ (Falk, Fehr, and Fischbacher, 2003; Falk and Fischbacher, 2006; Falk, Fehr, and Fischbacher, 2008).

Compared to these papers, we study consequentialist and deontological behavior in individual choice situations with neither strategic considerations nor any self-interest motive. Chen and Schonger (2022) also focus on situations without strategic interactions. They model deontological decision-making as governed by lexicographic preferences, first over acts *per se*, and secondarily over consequences. In two experiments, they show that when the consequences of a donation decision become increasingly hypothetical, subjects are more likely to donate money, again suggesting a mix of deontological and consequentialist motives. Bénabou and Tirole (2011) and Bénabou et al. (2023) show how behaviors displaying Kantian-like "sacred values" and "taboos tradeoffs" arise, even for consequentialist agents, in a model of moral identity maintained through self-signaling. In addition, a series of papers (Erat and Gneezy, 2012; Cappelen, Sørensen, and Tungodden, 2013; Gibson, Tanner, and Wagner, 2013; Biziou-van-Pol et al., 2015; Abeler, Nosenzo, and Raymond, 2019) provide evidence that people display an inherent aversion against lying even when doing so would generate Pareto improvements or morally desirable outcomes, which is consistent with deontological principles.

We add to this literature by providing evidence on the prevalence and, importantly, the consistency (or lack thereof) of consequentialist and deontological behavior across a range of decision situations. Furthermore, in contrast to the previous literature, we purposefully abstract from preferences based on self-other considerations. Apart from rule breaking, our ends-versus-means decision situations were designed so that subjects' self-interest plays *no role*, as their choices do not influence their own payoffs. This feature also allows us to investigate, in a second stage, the relationship between deontological/consequentialist motives and prosociality.

Our investigation and findings on deliberate randomization by subjects, indicative of preference uncertainty and ambiguity aversion, build on a recent literature on preferences for randomization (see Agranov and Ortoleva (2022) for an overview) and "cautious utility" as a potential source of such behavior (Cerreia-Vioglio, Dillenberger, and Ortoleva, 2024).

At the boundary with moral psychology and philosophy, we also contribute to the large literature that uses the trolley dilemma to investigate underlying moral principles (Greene et al., 2001; Hauser et al., 2007; Bartels, 2008; Lanteri, Chelini, and Rizzello, 2008; Rai and Holyoak, 2010; Bauman et al., 2014; Costa et al., 2014; Gawronski and Beer, 2017; Awad et al., 2020). These studies typically focus on moral judgments instead of behavior and use hypothetical situations and questionnaires. We instead focus on incentivized behavior, and further add to the literature a comprehensive assessment of the predictive power of the trolley dilemma for behavior in other EVM dilemmas, as well as in SVO tradeoffs.[4]

2 Conceptual framework

We provide here a simple framework that formalizes our research questions and experimental approach. We first define ends-versus-means (EVM) tradeoffs and preference specifications that capture consequentialist and deontological concerns. We then use the framework to analyze what can be inferred about such preferences from individuals' choices in EVM situations. Finally, we repeat these steps for preferences over self-versus-other tradeoffs.

2.1 EVM Tradeoffs, consequentialism, and deontologism

Choices. Agents face a decision task $t \in T$ in which they have to choose one of two actions, $a \in \{a_{1t}, a_{2t}\}$. Action a_{1t} generates a material payoff x_t for some other agent(s), while a_{2t} generates $X_t > x_t$ for (possibly different) other recipients. Note that neither action involves a gain or loss for the decision maker, in contrast to standard self-versus-other tradeoffs like dictator and public-goods games. At the same time, there is a rule r_t, such as some instantiation of the Kantian imperative or a religious commandment, that prescribes choosing one of the actions over the other. A choice task is thus defined by $(a_{1t}, a_{2t}, x_t, X_t, r_t)$: the two actions, their material consequences, and the rule involved. In any given choice situation, we will call the action prescribed by the rule the deontological action, denoted d, and the action yielding the higher social payoff the consequentialist action, denoted c.

[4]Three papers also examine the role of consequences in trolley dilemmas. Hsu, Anen, and Quartz (2008) and Gold, Pulford, and Colman (2015) create a trolley-like incentivized scenario in which subjects divert meals among orphans. They compare behavior and moral judgments, finding no difference. Bostyn, Sevenhant, and Roets (2018) design a scenario in which subjects believe they are distributing electrical shocks among mice. The experiment involves deception, as no shocks are actually delivered. They observe significantly fewer deontological choices relative to a fully hypothetical scenario. In contrast to these two papers, our SAL-trolley paradigm involves actual human lives (albeit statistical ones), and is thus closest to the classical version of the dilemma.

Ends-versus-means tradeoffs are situations in which achieving the best social consequences requires taking actions that are themselves morally objectionable or impermissible, so that $c \neq d$. For instance, in the classical trolley dilemma, one must choose between diverting a runaway trolley onto a track where it will kill one person, or leaving it on its current path where it will kill five. Diverting is the consequentialist action, as it saves more lives, but violates the commandment against actively causing someone's death, as well as Kant's imperative never to treat a human being solely as a means to an end (Kant, 1785). Thus, not diverting the trolley is the deontological option.

Preferences. Consider an individual i facing a task t with $c \neq d$ and denote by V_t^i and $v_t^i < V_t^i$ the utilities they derive when other agents receive payoffs X_t or x_t, respectively. Let $r(a)$ be an indicator equal to 1 if and only if the chosen action respects the rule, i.e., if $a = d$. An agent with purely consequentialist concerns has utility

$$U_{c,t}^i(a) = r(a)v_t^i + (1 - r(a))V_t^i,$$

and thus chooses $a = c$ in every task, ignoring the rule. Conversely, an agent with purely deontological concerns has utility

$$U_{d,t}^i(a) = r(a)k_t^i - (1 - r(a))K_t^i,$$

where $k_t^i > 0$ is the psychological payoff from choosing the action that conforms to the rule r_t and $-K_t^i < 0$ is the psychological loss from violating it. Such agents choose $a = d$ in every task irrespective of consequences.

Because such extreme preferences may be unrealistic, we next allow agents to have both deontological and consequentialist concerns, with weights $(\alpha^i, 1 - \alpha^i)$ that are stable across tasks. An agent with such *mixed preferences* has utility

$$U_t^i(a) = \alpha^i U_{c,t}^i(a) + (1 - \alpha^i)U_{d,t}^i(a), \tag{1}$$

in which rule violations are penalized and weighted against material consequences. The agent then chooses $a = d$ in task t if and only if:

$$\frac{\alpha^i}{1 - \alpha^i} > \frac{V_t^i - v_t^i}{K_t^i + k_t^i} \equiv \frac{1}{z_t^i}. \tag{2}$$

Finally, because moral choices can be difficult, agents may be unsure about which principle they ought to follow and afraid of making the "wrong decision," leading to regret or guilt later on. We model this *normative uncertainty* by allowing an agent to be uncertain about the extent to which they have consequentialist or deontological preferences –a form of imperfect self-knowledge– and averse to such ambiguity, in the spirit of "cautious utility" (Maccheroni, 2002; Cerreia-Vioglio, Dillenberger, and Ortoleva, 2024), but allowing for smooth ambiguity aversion (Klibanoff, Marinacci, and Mukerji, 2005). Thus, for agent i in task t, with some probability $0 \leq \omega \leq 1$, they have deontological preference $U_{d,t}^i(a)$, and with probability $1 - \omega$ consequentialist preference $U_{c,t}^i(a)$, where ω is unknown to the agent and follows some distribution $F^i(\omega)$. They make choices by maximizing

$$U_t^i(a) \equiv \int_0^1 \Psi_i[\omega U_{c,t}^i(a) + (1-\omega)U_{d,t}^i(a)]\, dF^i(\omega). \tag{3}$$

where Ψ is a concave function, assumed here to be $\Psi_i(u) = -e^{\eta^i u}$, where $\eta^i > 0$ parametrizes ambiguity aversion. For simplicity and easy comparability with the previous cases, we will take the distribution $F^i(\omega)$ to be binary, with weights $(\alpha^i, 1-\alpha^i)$ on 0 and 1 (pure preferences) respectively.[5]

An agent with such uncertain moral values may want to hedge against ambiguity by randomizing their choices.[6] With binary tasks, each performed only once, we observe only the outcome and therefore cannot detect randomization. However, if given the option to make an explicit, continuous probabilistic choice, agent i will select $a = d$ in task t with probability

$$q_t^i = 1 - \frac{\ln\left(\dfrac{1-\alpha^i}{\alpha^i} \cdot \dfrac{V_t^i - v_t^i}{K_t^i + k_t^i}\right) - \eta^i\,(v_t^i - k_t^i)}{\eta^i\,(V_t^i - v_t^i + K_t^i + k_t^i)} \in (0,1), \tag{4}$$

provided that

$$\eta^i\,(v_t^i - k_t^i) \;<\; \ln\left(\frac{1-\alpha^i}{\alpha^i} \cdot \frac{V_t^i - v_t^i}{K_t^i + k_t^i}\right) \;<\; \eta^i\,(V_t^i + K_t^i), \tag{5}$$

Quite intuitively, q_t^i is increasing in the internal uncertainty and preference parameters $(\alpha^i; \eta^i, K_i^t, v_t^i)$ that are relevant to the the worst-case scenarios: choosing $a = c$ when (upon further reflection) one is really a deontologist, or choosing $a = d$ when one is really a consequentialist. In contrast, the comparative statics in V_t^i and k_t^i are ambiguous, a type of non-monotonicity frequent in ambiguity aversion models.

2.2 Testing for pure and mixed preferences

Our first step will be to examine, in a series of incentivized ends-versus-means choice tasks, the overall extent of consequentialist versus deontological decisions.

Test 1 (Prevalence of consequentialist and deontological concerns)**.** *If c and d are both chosen in significant proportions, we can reject the standard assumption that all agents are consequentialist, as well as all being purely deontological. If a significant fraction of subjects make varying choices across tasks, the population cannot be a mix of pure moral types.*

Since agents' choices are unlikely to satisfy the invariance requirement implied by pure types, a more meaningful test of consistency allows for preference types that put weight on both moral concerns. Generating regularities in behavior now requires more structure on the utility payoffs $(V_t^i, v_t^i, K_t^i, k_t^i)$, however.

[5]Equivalently, ω is the unknown weight in a mixed preference. The case (1) corresponds to a distribution $F^i(\omega)$ that reduces to an atom at α^i, or alternatively to the limiting case $\eta^i = 0$ (ambiguity neutrality), with $\alpha^i = \int_0^1 \omega dF_i(\omega)$. Note that, with $\eta \to 0$, (5) gives back the threshold behavior in (2).

[6]For a survey of the evidence on deliberate randomization, see Agranov and Ortoleva (2022).

To this end, let us make two assumptions on the relative preference intensity defined in (2), which we state intuitively here, then formally below: (i) individuals with strong deontological concerns α_i also tend to evaluate each task's ratio z_t^i of rule-breaking losses to consequential gains as high; (ii) for any individual, if this ratio is high in some one task (say, the trolley), it also tends to be high in any other task (say, the lying game). Under these assumptions, an agent's weight α_i defines a type that generates consistent behavior across tasks.

Consistent mixed preferences. Let preferences be given by (1) and assume that, for each i, the random variables $(\alpha^i, \{z_t^i\}_{t\in T})$ are affiliated.[7] Then, choices differ across both subjects and tasks, but:

1. Individuals' decisions (and their probabilities of choosing d) are positively correlated across tasks: $\mathrm{Cov}(\mathbb{1}_{a_t^i=d}, \mathbb{1}_{a_{t'}^i=d}) > 0$, or equivalently, for each i, $\Pr[a_{t'}^i = d | a_t^i = d] > \Pr[a_{t'}^i = d]$ for all t, t'.
2. Furthermore, if in any task t subject i chooses $a_t^i = d$ while subject j chooses $a_t^j = c$, the posterior belief over α_i first-order stochastically dominates that over α_j, and hence subject i is more likely than subject j to choose d in any other task t'.

Such consistency, or lack thereof, is the second question we study in our main experiment:

Test 2 (Consistency of consequentialist and deontological concerns)**.** *We can assess the extent to which subjects have consistent propensities to act consequentially or deontologically by examining whether there is a significant correlation of decisions across EVM tasks.*

2.3 Testing for normative uncertainty and cautious preferences

Another reason why agents' choices may vary across tasks is if they deliberately randomize their decisions, as predicted by ambiguity aversion under preference uncertainty. Importantly, we would also like to know *per se* whether, in some dilemma situations, agents indeed experience normative uncertainty and hedge their choices as a result.

Test 3 (Prevalence of normative uncertainty and cautious preferences)**.** *We can assess the prevalence of normative uncertainty and cautious preferences by examining the extent to which subjects randomize their choices when given the opportunity to do so explicitly.*[8]

If there is indeed randomization, we can next examine the consistency of this behavior across tasks. Stable preference types correspond, intuitively, to probabilities q_t^i's that are positively correlated

[7]A set of random variables $\{X_1, X_2, \ldots, X_n\}$ that are affiliated (Milgrom and Weber, 1982) are also *associated,* meaning that, for any two non-decreasing functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$, $\mathrm{Cov}[f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)] \geq 0$. The simplest case is here when the z_t^i's are independent of α^i and i.i.d. In our four donation-based EVM tasks, the payoffs X_t, x_t are the same, which is most likely a further source of positive covariation in the $V_t^i - v_t^i$'s, hence in the z_t^i's. If the (K_t^i, k_t^i)'s vary over agents but not tasks, an individual's composite degree of deontologism is $\alpha^i(K^i + k^i)/(1 - \alpha^i)$.

[8]At a general level, deliberate randomization only reveals convexity in preferences. Its most natural source in our setting, however, is ambiguity aversion over what one's moral preferences really are, or should be. That is therefore how we shall interpret it henceforth.

across tasks and tend to be systematically higher for some individuals than for others. Recalling that q_t^i is increasing in $(\eta_i, \alpha^i, K_t^i; v_t^i)$ but non-monotonic in k_t^i and V_t^i, we will focus on the most intuitive case, where an agent i's type is defined by a task-invariant triplet (α^i, η^i, K^i).

Consistent cautious preferences. *Let preferences be given by* (3) *and assume that: (i) for each* i*,* (α^i, η^i, K^i) *are task-invariant and affiliated; (ii)* $\{(V_t^i, v_t^i, k_t^i)\}_{t\in T}$ *are independent of* (α^i, η^i, K^i) *and iid. Then:*

1. *Individuals' randomization probabilities (and therefore their decision outcomes) are positively correlated across tasks:* $\mathrm{Cov}(q_t^i, q_{t'}^i) > 0$*.*

2. *If in any task* t *subject* i *chooses* $a_t^i = d$ *while subject* j *chooses* $a_t^j = c$*, the posterior belief over* (α^i, η^i, K^i) *first-order stochastically dominates that over* (α^j, η^j, K^j)*, and hence subject* i *is more likely than subject* j *to also choose* d *in any other task* t'*.*

As noted, correlation in choice probabilities implies correlation in realized choices. However, since probabilities are the relevant strategies here, and we are also interested in preference uncertainty itself, it is more natural to look directly at those. Moreover, in practice, subjects may find it easier (lower cognitive costs) to randomize when offered a user-friendly device to do so than when they need to do it "in their heads", which is the case when they must choose either c or d.

Test 4 (Consistency of normative uncertainty and cautious preferences)**.** *We can assess the extent to which subjects' uncertain moral identity and cautious preferences are consistent by examining whether their randomization probabilities are significantly correlated across tasks.*

2.4 Self-versus-other tradeoffs

In contrast to EVM tradeoffs, where actions solely affect others' payoffs, in SVO tradeoffs they also affect one's own payoff. In Appendix A, we first define pure selfish and altruistic types, who differ in how they value others' versus their own payoffs and in how much negative utility (e.g., guilt) they experience from acting selfishly. We then define mixed types and types with uncertainty over their own prosociality just as before. Our tests for the prevalence and consistency of selfish, prosocial, mixed and cautious preferences then follow in the same way. We thus first examine the frequency of each type of choice and their correlation across an individual's decisions. We then assess how often subjects randomize, and how correlated the chosen probabilities are across tasks. Finally, we will examine whether decisions (and probabilities) in SVO choice tasks are correlated with those in EVM choice tasks.

3 Experimental design

Our main design consists of two main blocks of choice tasks, followed by questionnaires. The first one confronts subjects with a set of "ends-versus-means" (EVM) decisions. The second uses well-established "self-versus-other" (SVO) decisions.

3.1 Ends-versus-means tradeoffs

3.1.1 Description and motivation

Table 1: Ends-versus-means choice tasks

Choice task	Ends (Payoffs X_t and x_t)	Means (Rule r_t)
Trolley dilemma	Life of three people vs. one person	Sacrifice vs. spare one person
Lying game	Donation vs. money to other subject	Lie vs. tell truth to another subject
Bribe game	Donation vs. money to other subject	Bribe vs. not bribe another subject
Group donation game	Donation vs. money to other subject	Consequential vs. inconsequential voting
Statement choice	Donation vs. money to other subject	Make vs. not make repugnant statement
Rule-following task	No money vs. money to self	Follow vs. break rule

We designed six EVM choice tasks, summarized in Table 1. We provide here an overview of each one, which is a direct counterpart to one of Kant's instantiations of his Categorical Imperative. A detailed description of each task is provided in Appendix B and the instructions to subjects are in Appendix J.

Trolley dilemma. *"Even if by the death of one innocent man the whole world could be preserved, one ought not to kill him."* (Kant, 1788)

The trolley dilemma is one of the most popular thought experiments in moral philosophy and applied ethics, used extensively to distinguish between modes of moral decision-making. It is, however, frequently criticized for representing an artificial and implausible situation, detached from any decisions people face in reality (e.g., Bauman et al., 2014). We implement the trolley dilemma with real, meaningful consequences using the *Saving a Life* Paradigm (Falk and Graeber, 2020). Key features of our version are that choices involve the saving of actual human lives, and that they occur in a realistic context that nonetheless contains all relevant features of the trolley dilemma.

We partner with a charity that treats people in India suffering from Tuberculosis. Subjects face a setting in which a donation of 380 Euros to the charity is preset, allowing it to treat five patients in some Indian state. Based on the medical literature, this saves, in expectation, one life that would otherwise have been lost to the disease. Subjects can allow this default donation to occur or redirect it to some other state, in which case it is tripled, so that fifteen people will be treated. By doing so, they forego saving one life in order to save three elsewhere, in expectation – a real trolley dilemma. Our decision interface is also designed to closely mimic common representations of the classical trolley, with redirection achieved by pulling a slider on the screen. See Appendix Figure B.1 for a screenshot and further details.

Core EVM tasks. Following the SAL trolley, we designed four core EVM choice tasks in which we fix the social payoffs $X_t = X$ and $x_t = x$ to actions c and d and only vary the rule r_t that prescribes choosing d. Specifically, X means that a charity supporting children suffering from cancer will receive a donation of 15 Euros, while x means that another randomly selected subject in the experiment will receive 2 Euros. As in the trolley, there is no self-versus-other tradeoff, a feature that fundamentally distinguishes these games from standard ones studied in the literature on social preferences. We chose the two social consequences X, x so that the associated utilities will be ordered as $V_t^i > v_t^i$ for almost everyone. Not only does the donation involve substantially more money, but helping children in need plausibly generates higher utility than increasing the payoff of another subject. Accordingly, in the absence of other rules/means, both deontological and consequentialist reasoning plausibly agree that choosing the donation is the morally right choice. To verify that subjects in our experiment also think this way, we gave them the simple choice between the two payoffs, and indeed 93% of them chose the donation.[9] Having established a setting in which the morally desirable consequence is largely undisputed by subjects themselves, we then systematically vary the means involved to bring about these ends.

Lying game. *"Truthfulness in statements that cannot be avoided is the formal duty of a human being to everyone, however great may be the disadvantage that may arise therefrom."* (Kant, 1797b)

Subjects play a sender-receiver game, adapted from Gneezy et al. (2020) into an EVM tradeoff. The receiver must choose between two options, A and B, not knowing anything about their consequences except what the sender tells them. The sender knows that A will trigger the 15 Euro donation, while B will increase the payoff of the receiver by 2 Euros. Our focus is on the behavior of the senders, who can communicate one of two messages. They can either tell the receiver that Option A will give them the higher payoff- a lie - or that Option B will give them the higher payoff - the truth. As almost all receivers follow the message (which we empirically verify), senders face a tradeoff between lying to trigger the donation and telling the truth.

Bribe game. *"He who by gifts seeks to purchase another's decision acts contrary to duty; for he intends to make use of the other merely as a means to his own end."* (Kant, 1797a)

There are a sender and receiver, each now controlling some donation. Our focus is again on senders, who decide whether to implement the 15 Euro donation or increase the receiver's payoff by 2 Euros. Receivers can then choose between implementing a larger 20 Euro donation and increasing their own payoff by 2 Euros. Using the strategy method, we elicit their conditional response to each possible choice of the sender and report it to the sender before they make their decision. Importantly, we match each sender with a receiver who has precommitted to choosing the 20 Euro donation *only if* the sender sent them 2 Euros rather than donate. In effect, they demand a bribe in order to implement the more socially desirable option (the word bribe is never used in the instructions). Thus, senders do not face a strategic situation but a tradeoff between sending 15 Euros to the charity without regard to ultimate consequences and raising the total donation to 20 Euros by paying the required "bribe."

Group donation game. *"Act only according to that maxim whereby you can at the same time will that it should become a universal law."* (Kant, 1785)

In this game, adapted from Falk, Neuber, and Szech (2020) into an EVM tradeoff, subjects are placed into groups of six that are collectively entrusted with the 15 Euro donation. The first five each decide simultaneously between preserving this donation and taking 2 Euros for themselves. If any of them makes the latter choice, the group's donation is canceled. Afterwards, the sixth group member

[9] Furthermore, when asked, 91% of subjects agree that from a moral point of view, there is a right and a wrong decision for them in this situation; and of those, 98% name choosing the donation as the right decision. Similarly, 95% think that other people consider the donation to be the morally right choice.

–our subject of interest– must state whether they want to implement the donation or send the usual 2 Euros to another (non-group) participant. Importantly, before deciding, they learn whether the donation has already been canceled by the decisions of the first movers, which in our experiment happened in every group. The sixth movers thus face a tradeoff between an expressive but totally inconsequential "vote" in favor of a donation that will not be implemented in any case and the consequentialist option that at least benefits some other participant.

Statement choice. *(Relevant to both the categorical imperative's universalization principle and its proscription against lying, cited above)*.

Subjects decide whether to make a morally repugnant statement (supporting the destruction of the environment) or a positive one (supporting the preservation and protection of the environment). Making the repugnant statement triggers the 15 Euros charitable donation with probability 50%, while the virtuous one triggers the 2 usual Euros for some other subject. Neither has any other consequence, including on the environment. In particular, choices are anonymous (as always) and subjects are informed that the statements will not be used for any purpose, such as an opinion poll, thus depriving those of any instrumental value. Subjects thus face a tradeoff between making the repugnant statement to achieve the 15 Euro donation and making the agreeable statement, thereby sticking to their personal values and generating a small social benefit on top of that.[10]

In addition to the SAL trolley and our four core EVM choice tasks, we test subjects' willingness to comply with rules of proper conduct out of their own volition.

Rule-following task. Subjects guide a stick figure across a sequence of red traffic lights on the screen (adapted from Kimbrough and Vostroknutov (2016)). They are told that the rule is to wait at each light until it turns green, which happens after five seconds, but there is no enforcement of any kind. Waiting is costly, as they are endowed with 8 Euros and lose 8 cents each second until the end is reached. Subjects can break the rule, however: at each traffic light, they can press a button to "run the light" and proceed immediately. They thus face a tradeoff between following the non-instrumental rule and breaking it to maximize earnings.[11]

3.1.2 Types of treatments

We implemented three treatments, each featuring different variants based of the EVM choice tasks and conducted as a separate experiment.

EVM (ends-versus-means) decisions. These main conditions feature the EVM tradeoffs as described in the previous section. That is, in each one there is a moral conflict between the consequences of the choice options and the actions required to bring them about.

[10]Our sample consists of very environmentally-conscious subjects: 96% agree that fighting climate change is important, and 99% that measures to protect the environment are important. For the link between Kantian-like behaviors and moral-identity maintenance, see Bénabou and Tirole (2011) and Bénabou et al. (2023). Bursztyn et al. (2020) show that many Pakistani men forego significant transfers when they have to express gratitude to the US government for providing it; this case involves a payoff to oneself, unlike our EVM tradeoffs.

[11]In order to stay close to the original design, breaking the rule is costly to the subjects themselves. In this case there are consequences to the subjects themselves, but as there are no consequences to others and thus no self-versus-other tradeoff, this detail does not impede the ends-versus-means comparison.

EMA (ends-means-aligned) decisions. To test for the robustness of our results against concerns such as limited attention or comprehension, we designed counterparts to the main tasks but in which ends and means are now *aligned*. Thus, in the trolley the default is now a donation that will save five people, while redirecting it by pulling the slider divides it by three, saving only one person. In the lying game, it is now the truthful message that triggers the charitable donation; in the statement game, it is endorsing the virtuous rather than repugnant assertion, and so on. Deontological and consequentialist principles now prescribe the same option, while other features of each game, and in particular its complexity, remain invariant. Comparing subjects' behavior between the original and aligned versions allows us to check whether they pay attention to, and properly understand, the decision problems they face.

EVM-EMA decisions with randomization option. We designed variants of both the EVM and the EMA tasks that allow subjects to randomize between options. Following Agranov and Ortoleva (2017) and Agranov and Ortoleva (2023), they choose an integer between 0 and 10, corresponding to a probability $(0 = 0\%, 1 = 10\%, \dots)$ according to which options are enacted (if that decision is selected for implementation). For example, in the trolley dilemma, a choice of $a = 7$ implies a 70% chance that the donation is not redirected and a 30% chance that it is, and so on. We created the option to randomize for all EVM choice tasks with the exception of the rule-following game. This one was excluded due to its animation mechanics, which prevent implementing the same randomization option as in the other choice tasks.

Robustness treatments. As described in Section 3.4 below, we included extensive comprehension checks in all experimental conditions. For the trolley dilemma, we also included two further robustness treatments. Our SAL version of the dilemma differs from moral philosophers' classical one in two ways. First, the scenario is not about imagined train tracks and repairmen happening to be on them, but about actual patients and treatments for a real and common disease. To investigate the effect of this added realism, subjects also faced the classical train-track trolley version. Second, subjects' choices are not hypothetical but can have real, important consequences. To properly isolate the effect of this latter feature, we ran two different between-subject treatments. In Treatment *SAL-Hypothetical*, the choice environment is exactly as described above, but all choices are, and are presented as, hypothetical: there is never any actual donation. In contrast, in Treatment *SAL-Real*, for each subject there is a 10% probability that their decision will be implemented, resulting in either one life saved in State A or three in State B.[12]

3.2 Self-versus-other tradeofs

3.2.1 Description and motivation

In addition to a systematic investigation of consequentialist versus deontological decision-making, we are also interested in its potential relation to the more standard dimension of morality, namely selfishness versus prosociality. Our second block of choice tasks therefore consists of standard games

[12]We ended up implementing the decisions of 23 subjects. As a result of the ensuing donations in both states, 265 patients were treated, and thus 53 lives were saved, in expectation.

from the literature on social preferences, listed in Table 2. Their defining feature is that subjects now face tradeoffs between their own monetary outcomes and those of other subjects, so we label these situations self-versus-other (SVO) tradeoffs. Most are familiar, and we implement them using standard procedures. Therefore, we leave the description of procedural details to Section C of the Appendix and provide here only a brief summary.

We include the following choice tasks. (i) A standard *dictator game*, in which subjects are endowed with 20 Euros and decide how to allocate it between themselves and the children's charity. (ii) A modified *dictator game with taking*, in which the charity is endowed with the 20 Euros and the subject can take money away from it to increase their own payoff. (iii) A standard *trust game*, which subjects play in both the sender and receiver roles. As sender, they decide how much of an endowment of 5 Euros to send to another subject; this is then tripled. As receiver, they decide, using the strategy method, how much to return for all possible amounts sent. From these decisions we construct three measures: (iii) the amount sent, as measure of *trust*, (iv) the amount sent back when having received 1 Euro, measuring *Positive Reciprocity* at a *Low* send amount, and (v) the amount sent back when having received 5 Euros, measuring *Positive Reciprocity* at a *High* send amount. (vi) Subjects also play a standard public-goods game with two other participants, each deciding how much to contribute from their 5 Euro endowment. (vii) Lastly, we include a *Moral luck* game, in which subjects acting as dictators decide how much of their endowment to give to another subject, based on the latter's previous behavior. Specifically, the other subject had the choice between two lotteries, with 70%-30% and 30%-70% chances of benefiting themselves or the charity, respectively. This generates four choice-outcome cases: the charity receives a donation, either intended or not intended by the subject; or the subject receives money, either intended or not intended. How much dictators give in each case allows us to measure the extent to which they reward: (vii) intentions, namely the other subject's *ex-ante* lottery choice; (viii) consequences, namely, its *ex-post* realization; (ix) one versus the other.

3.2.2 Types of treatments

As for the EVM tradeoffs, we also implemented a version of the SVO choice tasks with a randomization option. To do so, we first binarized the games, then added the same integer choice between 0 and 10 defining implementation probabilities for the two options. For example, in our dictator game, either the entire endowment is allocated to the subject, or it is equally split between them and charity. Then, if a subject chooses $a = 9$, the equal split is implemented with 90% probability, while with 10% probability they receive the entire endowment, and so on. We included the option to randomize for all SVO choice tasks except for the moral-luck game, due again to implementation difficulties.

3.3 Questionnaires

To complement our incentivized behavioral tasks, we administered several widely used questionnaires, which constitute the third block of our design. First, we use the Oxford Utilitarianism Scale (OUS, Kahane et al., 2018), which elicits subjects' "permissive attitudes toward instrumental harm", much like in the trolley dilemma (OUS-IH), and their "impartial concern for the greater good", a

Table 2: Self-versus-other choice tasks

Variable	Elicitation method	Definition
Altruism	Dictator game with charity as the recipient, giving frame, 20 Euros endowment	Amount allocated to charity
Altruism Taking	Dictator game with charity as the recipient, taking frame, 20 Euros endowment	Amount allocated to charity
Trust	First mover in trust game, 5 Euros endowment	Amount sent to second mover
Pos. Recipr. Low	Second mover in trust game when having received 1 Euros out of 5 Euros (strategy method)	Amount send back to first mover
Pos. Recipr. High	Second mover in trust game when having received 5 Euros out of 5 Euros (strategy method)	Amount send back to first mover
Public goods game cont.	Public goods game, 5 Euros endowment, group size of 3, 0.5 marginal per capita return from contributing	Amount contributed
Rewarding intentions	Moral luck game	Amount allocated to first mover in $S4 - S3$
Rewarding cons.	Moral luck game	Amount allocated to first mover in $S4 - S2$
Rewarding cons. - intent.	Moral luck game	Amount allocated to first mover in $(S4 + S3) - (S2 + S1) - ((S4 - S3) + (S2 - S1))$

Notes: For details on the choice tasks, see Appendix C.

utilitarian-like tendency to judge the well-being of every individual as equally important (OUS-IB). Second, we administer the Moral Foundations Questionnaire (MFQ, Graham et al., 2011), which aims to measure five distinct dimensions of people's moral concerns: care versus harm of others, fairness/reciprocity, in-group/loyalty, authority/respect, and purity/sanctity. Third, we include the Moral Universalism short module of Enke, Rodríguez-Padilla, and Zimmermann (2022). Using hypothetical allocation games, it measures the extent to which subjects exhibit the same level of altruism towards strangers as towards in-group members (MU scale). Finally, we confront subjects with the classical, hypothetical train-track trolley dilemma, and also include a module on political attitudes and religiosity. For sociodemographic variables, we collected age, gender, subject of studies, final high-school grade, and gross monthly income.

3.4 The experiments

We conducted three experiments: *Main*, *Aligned*, and *Randomization*, with a total of 1,239 subjects. Each was implemented online as a virtual-lab experiment using oTree (Chen, Schonger, and Wickens, 2016), with the subject pool of the BonnEconLab. Subjects were invited using hroot (Bock, Baetge, and Nicklisch, 2014) and had to log in at a specified date and time, with an experimenter available throughout to answer questions and address any issues. Subjects earned 12 Euros as show-up fee.

Main Experiment. Subjects had to complete two sessions, separated by one week, each lasting about 45 minutes. One session contained the EVM choice tasks and the other one the SVO situations and the questionnaires. The order of sessions, as well as the order of decision tasks within each session, was randomized. In each session, one decision was selected for real implementation. In addition, in the *Real* treatment of the SAL trolley, every subject's choice was implemented with 10% probability.

Overall, 626 subjects took part in the main experiment, of which 593 completed both sessions. Based on the pre-registration, we excluded the top 1% fastest subjects, as well as those who preferred giving 2 Euros to another participant over a 15 Euros donation to the children-with-cancer charity in the baseline task. The reason for excluding them is that no opposing predictions in the ends-versus-means games exist for them. The final sample thus consisted of 548 subjects (339 women, mean age = 27, SD = 8). The results fully replicate when using the full sample instead of the pre-registered one.

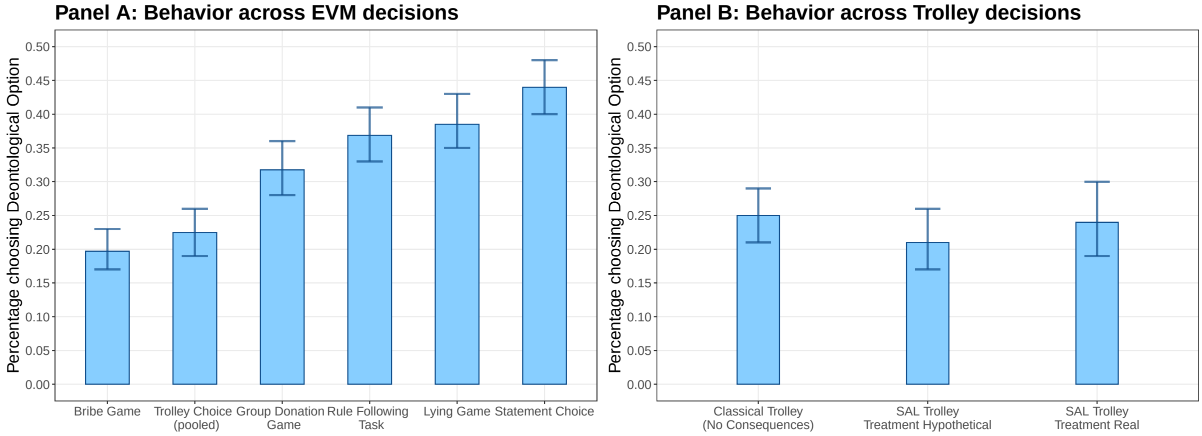
Aligned Experiment. Subjects faced the EMA choice tasks in a single session. In total, 135 subjects participated. After applying the same preregistered exclusion criterion as described above, we were left with 122 subjects (82 women, mean age 24, SD = 5). As in the main experiment, we randomized the order of decision tasks, and one decision in each session was selected for real implementation, excluding the aligned version of the trolley dilemma, which was stated hypothetically.

Randomization Experiment. Subjects randomly faced one of three decisions: the *EVM*, *EMA*, or *SVO* decisions, each with the option to randomize, and they completed the experiment in a single session. In total, 478 subjects participated. After applying the same exclusion criterion, we were left with 443 individuals (272 women, mean age = 27, SD = 8). Results remain robust when considering the full sample. In total, 152 subjects faced the EVM, 151 the EMA, and 140 the SVO decisions. All other aspects of the experiment were implemented as in the main experiment.

Comprehension. In order to ensure that all participants fully understood the decision rules and consequences, we included extensive comprehension checks throughout all the experiments. For instance, each of the major choice tasks (except rule-following and statement choice) featured a quiz after its introduction and subjects were only allowed to proceed once they answered all questions correctly.[13] Moreover, in the main EVM experiment, we also implemented attention and memory checks for the group-donation game and the statement choice. At the end of the experiment (approximately 15-25 minutes after their decisions), subjects were asked, unannounced: (i) in the group-donation game, whether the donation was destroyed by the choices of the other group members; (ii) in the statement choice, what were the consequences of making the repugnant statement. Proper recall at that later time constitutes a lower bound on comprehension and attentiveness during the choice phase, allowing us to verify the robustness of the results.

[13]Upon giving a wrong answer in the comprehension questions, subjects were informed that one or more of their choices was incorrect and that they needed to check the instructions again.

Figure 1: Prevalence of deontological decision-making across decisions and contexts



Notes: Each bar displays the fraction of subjects choosing the deontological option - i.e., not pulling the lever - in the trolley dilemma. The first two bars display behavior in the classical train-track version, the last two behavior in the Saving A Life paradigm. Treatment *Real* is that in which the SAL trolley has real consequences, whereas in *Hypothetical* decisions are without consequences. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

4 Consequentialist and deontological decision-making

4.1 Prevalence

Test 1. Panel A of Figure 1, which displays the fraction of subjects who choose the deontological option in each EVM game, shows two main results. First, in every case, this fraction represents a substantial minority and is statistically different from zero ($p < 0.001$, one-sample test of proportions). Second, it varies substantially across games. At the lower end, 20% of subjects refuse to “bribe” the receiver to achieve the higher donation, and in the trolley dilemma 22% decide not to redirect even though it would save three lives rather than one. The fraction goes up to 32% who choose the inconsequential option in the group donation game, 37% who follow the rule to wait at red lights, 39% who refuse to lie to bring about the charitable donation, and a striking 44% who are unwilling to make the repugnant statement. Out of the 15 possible pairwise comparisons between these fractions, 9 are significantly different from each other at the 5% level. This variation across choice tasks reveals that pure types, who either always follow the rule or always maximize welfare consequences, are rare in the subject population.

In the trolley dilemma, we can examine the prevalence of deontological versus consequentialist decisions in three different treatments. As shown in Panel B of Figure 1, the fraction of subjects choosing not to pull the slider is 24% in the *Real* version of our SAL-trolley game, 21% in the *Hypothetical* version, and 25% in the classical train-track dilemma.[14] These fractions are not statistically different from each other ($p = 0.48$ and $p = 0.24$ respectively) and always statistically different from zero ($p < 0.001$). The numbers also resemble those from other studies that have studied the classical (hypothetical) version of the trolley dilemma. For instance, in Awad et al. (2020), who study 70,000

[14]The classical train-track trolley was included in both the *Hypothetical* and the *Real* treatment. Fractions are not statistically different from each other ($p = 0.82$, two-sample test of proportions), so we pool both.

participants across 42 countries, 19% choose the deontological option. Our findings thus show that responses in the trolley dilemma are not sensitive to whether it is stated abstractly or concretely (imagined train tracks versus actual tuberculosis treatments versus), nor to the presence of real incentives.[15] In addition, we find no significant association of trolley behavior with any demographic variable. More generally, none of the demographic variable in our dataset reliably predicts the extent of deontological and consequentialist decision-making across the EVM choice tasks, see Appendix Figure D.1 for details.

In what follows, we discuss various robustness checks to rule out the possibility that the results could be driven by subjects' confusion or inattention, or by choosing at random because they do not really care (or believe in) the consequences of their decisions.

Robustness: Comprehension checks. In our instructions, we extensively covered the implications of each situation and administered comprehension quizzes as well as memory checks. For the latter, we asked subjects at the end of the experiment to recall the consequences of their potential choices in group donation and statement-choice games –respectively, whether their group's donation had already been canceled by others' actions, and which of the two environmental statements would trigger a donation. The fractions remembering correctly were 81% and 80%, respectively, indicating that they understood the implications of their options at the time of decision-making. Similar fractions were able to correctly answer all comprehension questions for the other choice tasks on their first try and less than 5% needed more than one try. When we restrict the sample to only those subjects who answered all comprehension questions correctly on their first try and passed all the memory checks, we also find that a significant fraction chose the deontological option in every EVM choice task (Appendix Figure F.1).[16]

Robustness: *Aligned* treatment. If random choice out of not caring or mistakes were significant factors affecting subjects' behavior, we should observe similar results in the EVA tasks as in their EVM counterparts, because the instructions have the same structure and feature the same degree of complexity; only the consequences of the two options have been switched around, so that both moral principles now agree on what is the right thing to do ($c = d$). Instead, we observe in each case that almost all subjects choose the aligned option, ranging from 88% in the lying game to 96% in the statement choice situation (see Appendix Figure F.3). Moreover, there is no indication that the games for which we find a high degree of non-consequentialist choices in the EVM tasks are those where subjects choose the non-aligned option to a higher degree in the EMA counterpart. For instance, the statement choice has the highest degree of deontological choices (44%), yet also the highest share of aligned choices (96%).

[15]In the Appendix, we investigate whether this finding might mask heterogeneity in the types of subjects choosing either option in the different versions. For instance, it could be that male subjects (say) are more likely to choose the deontological option when consequences are hypothetical, but less likely once real consequences are introduced. However, we find no evidence of heterogeneous treatment effects across a wide range of characteristics; see Appendix Figure D.3 for details.

[16]Likewise, we replicate this finding among the full set of subjects without the preregistered exclusion restrictions, with fractions ranging from 20% to 45% (Appendix Figure F.2).

Robustness: Trolley treatments. For the trolley, we also used comprehension questions that quizzed subjects on the consequences of their actions. We find that both the observed differences between the classical and SAL versions of the trolley dilemma, and those between the *SAL-Real* and *SAL-Hypothetical* versions, become even smaller when we focus only on subjects who scored high on the SAL trolley comprehension test (Appendix Figure D.2). Another concern might be the fact that since only 1 in 10 subjects in *SAL-Real* had their choices implemented, this could have reduced attention to the incentives or may have led subjects to discount their relevance. To address this concern, in Appendix Section E, we report the results of additional robustness calculations We compare the results of Bénabou et al. (2023), who use the same SAL paradigm and 10% implementation probability but now with a self-versus-other tradeoff (taking money versus triggering the donation) to a similar treatment on a new sample where those choices are hypothetical. We find that facing real consequences substantially alters subjects' behavior, reducing prosocial giving by as much as 50% compared to a hypothetical scenario. These findings suggest that the absence of a real-incentives treatment effect in the trolley task is not due to a low implementation probability, but specific to its ends-versus-means nature where two moral principles conflict, unlike in SVO choices.

We summarize our findings with respect to the prevalence of consequentialist and deontological choices as follows:

Result 1. *The share of non-consequentialist decisions ranges between 20% and 44% across the EVM choice tasks, consistent with substantial deontological concerns among subjects. This behavior reflects deliberate decision-making, not confusion, inattention or indifference. Behavior in the trolley dilemmas is robust to our changes in framing and incentives.*

4.2 Consistency

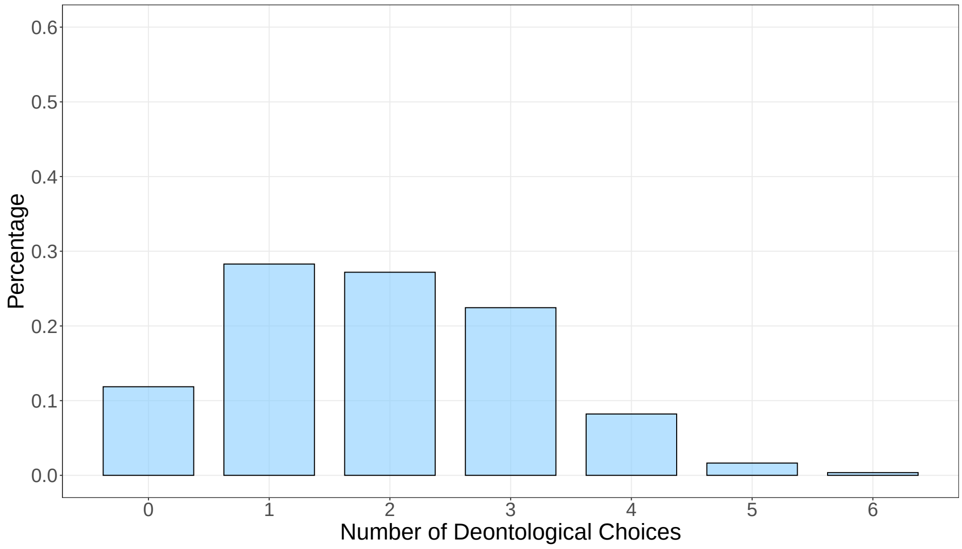
Having established the prevalence of both consequentialist and deontological concerns in EVM tradeoffs, we next investigate the extent to which they reflect stable preferences across decision settings, meaning that behavior in one such task is predictive of behavior in others.

Figure 2 provides a first graphical insight into the question by displaying the number of times a subject chose the deontological option. If pure types who behave consistently according to one of the two moral principles were frequent, we should observe a U-shaped distribution. Instead, the histogram displays an opposite, inverse-U shape. Essentially, no subject chooses the deontological option in all six cases, and only a minority of just over 10% behaves as standard economic agent, choosing in all six cases the consequentialist option.

Test 2. We now turn to our main test of consistency, allowing for mixed preferences, by analyzing the correlations of choices across EVM tasks. We start with the trolley dilemma due to its role as the principal thought experiment traditionally used to elicit moral intuitions about deontological versus consequentialist reasoning. Figure 3 displays, in pairwise flow charts, how choices in the SAL trolley relate to those in the other EVM games.[17] We find no evidence that choices in the former

[17]Because we neither find differences in the aggregate nor in individual characteristics between the *SAL-Real* and the *SAL-Hypothetical* treatments, we will pool decisions between treatments for this and all subsequent analyses. Our results remain unchanged if we use data from only one of the treatments.

Figure 2: Consistency of deontological and consequentialist decision-making



Notes: Histogram of the number of times a subject chose the deontological option in the ends-versus-means choice tasks.

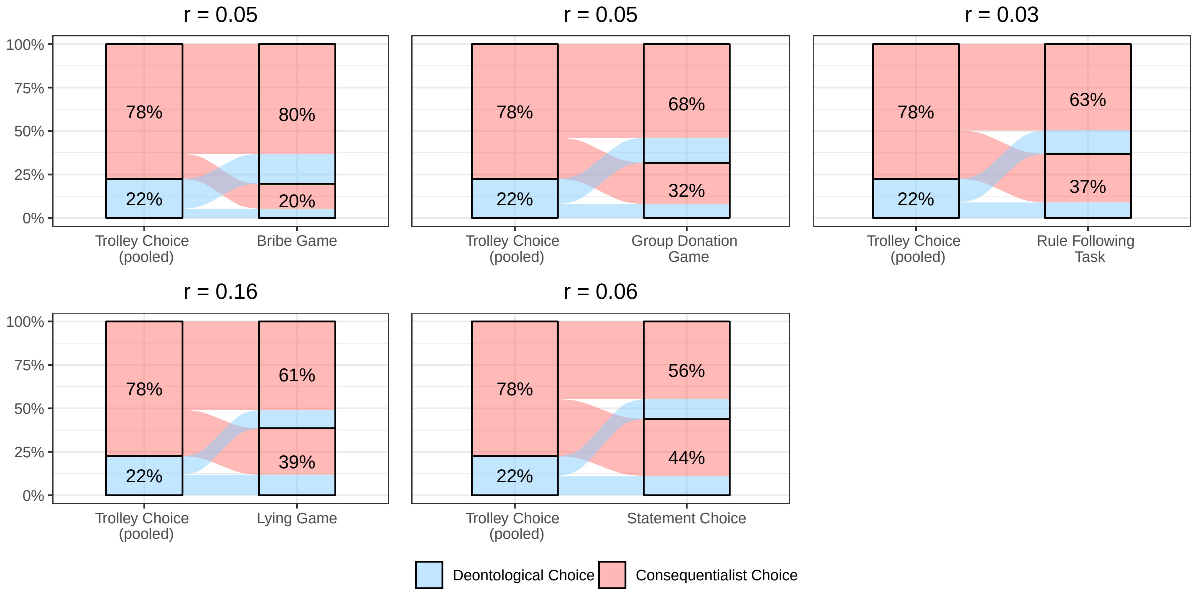
predict behavior in any of the latter. In each case, large fractions of subjects switch from choosing the deontological option in the trolley to the consequentialist one in the other situation, and vice versa. This is also reflected in the correlations between behaviors in the trolley and in the other EVM games (Pearson correlation tests), which are close to zero and insignificant ($p > 0.1$) for nearly all choice tasks. The only exception is the lying game, where a small but significant correlation does exist ($p < 0.01$).[18]

In Table 3, we extend the analysis to all pairwise correlations between the six EVM games. In each case, the indicator variable is equal to one if the subject chooses the deontological option (as defined in Section 3.1) and zero if they choose the consequentialist one. We find that no single choice task consistently predicts behavior in any other one: no correlation is larger than $r = 0.19$ and several are even negative. Using Cronbach's alpha as an aggregate measure of correlation, we find alpha $= 0.20$ (CI$=[0.09, 0.30]$), which is commonly considered very low. This lack of consistency means that individuals' choices do not seem to reflect preferences characterized by stable weights on consequentialist and deontological concerns. Instead, our evidence suggests that these relative concerns (e.g., the α^i's) are context-specific.

Robustness: comprehension. Could the lack of consistency be simply due to noise, because subjects are confused by or pay limited attention to our instructions? As we documented in the previous section, EVM behavior appears to reflect deliberate decision-making. To address any remaining concerns, we conduct three robustness checks for the analysis presented in Table 3. First, we exclude subjects who fail more than one of the comprehension quizzes (Appendix Table G.2). Second, we exclude those who fail one or both of the attention-memory checks we employed for the Group Donation and Statement Choice (Appendix Table G.3). Third, we exclude fastest 15% subjects (Appendix

[18]Note that these p-values are not corrected for multiple hypothesis testing. Applying such a correction would move the p-values even further away from significance.

Figure 3: Relationship of trolley behavior with behavior in the other ends-versus-means choice tasks



Notes: Each plot displays decision-making in the respective ends-versus-means game, conditional on behavior in the SAL-trolley dilemma. Percentages shown are the fractions of subjects choosing the deontological and consequentialist option in each case. Above each plot is the Pearson correlation coefficient between the SAL trolley and the respective choice task displayed.

Table G.4). In all three cases, we replicate almost zero correlations within the EVM block.

Robustness: trolley dilemma. Moreover, if confusion, inattention, or indifference was pervasive among our subjects and responsible for the lack of cross-task consistency, we would not expect any consistency between different framings of the same decision problem either. As a further robustness check, we therefore compute the correlations in subjects' choices between the SAL trolley dilemma and both the classical trolley dilemma and the instrumental-harm measure of the Oxford Utilitarian scale (OUS-IH). Across these dilemmas, the fundamental tradeoff stays essentially the same: whether to use means that cause instrumental harm to some people in order to achieve a greater overall good.

We find that these choices are highly correlated: the correlation of the SAL trolley with the classical one is $r = 0.50$ ($p < 0.001$), while its correlation with the OUS-IH is $r = 0.29$ ($p < 0.001$).[19] Furthermore, neither of those two predicts behavior in any other EVM choice tasks (correlations are never above $r = 0.15$), just as is the case for the SAL trolley. Together with the previous similarities in prevalence, these results show that subjects' behavior is stable across sacrificial dilemmas, but not across these and other EVM tradeoffs.[20]

We summarize our findings with respect to the consistency of consequentialist and deontological decision-making as follows:

[19]Replicating Kahane et al. (2018), we find a lower correlation of $r = 0.12$ of trolley behavior with the measure capturing subjects' instrumental benefit concerns (OUS-IB), which features a distinct EVM tradeoff.

[20]We also find no association between EVM behavior and any measure from the Moral Foundations Questionnaire (MFQ) or Moral Universalism (MU) scales. These scales appear to measure largely orthogonal dimensions of moral preferences or intuitions. For details, see Appendix Table G.1.

Table 3: Correlation matrix for ends-versus-means choice tasks

	Group Don. Game	Bribe Game	Statement Choice	Lying Game	Rule foll. Task
SAL Trolley	0.05 [-0.04, 0.13]	0.05 [-0.03, 0.14]	0.06 [-0.02, 0.14]	0.16 [0.08, 0.24]	0.03 [-0.05, 0.12]
Group Don. Game		0.14 [0.06, 0.23]	0.15 [0.07, 0.23]	-0.09 [-0.17, 0.00]	-0.06 [-0.14, 0.03]
Bribe Game			0.19 [0.11, 0.27]	0.04 [-0.04, 0.12]	-0.08 [-0.17, 0.00]
Statement Choice				-0.06 [-0.14, 0.02]	0.04 [-0.04, 0.12]
Lying Game					0.00 [-0.08, 0.09]

Notes: The table displays Pearson correlation coefficients between the ends-versus-means choice tasks. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

Result 2. *Behavior in one EVM task is not predictive of behavior in any other. Thus, we find no evidence of stable preference types placing task-invariant weights on consequentialist and deontological concerns.*

4.3 Normative uncertainty and cautiousness

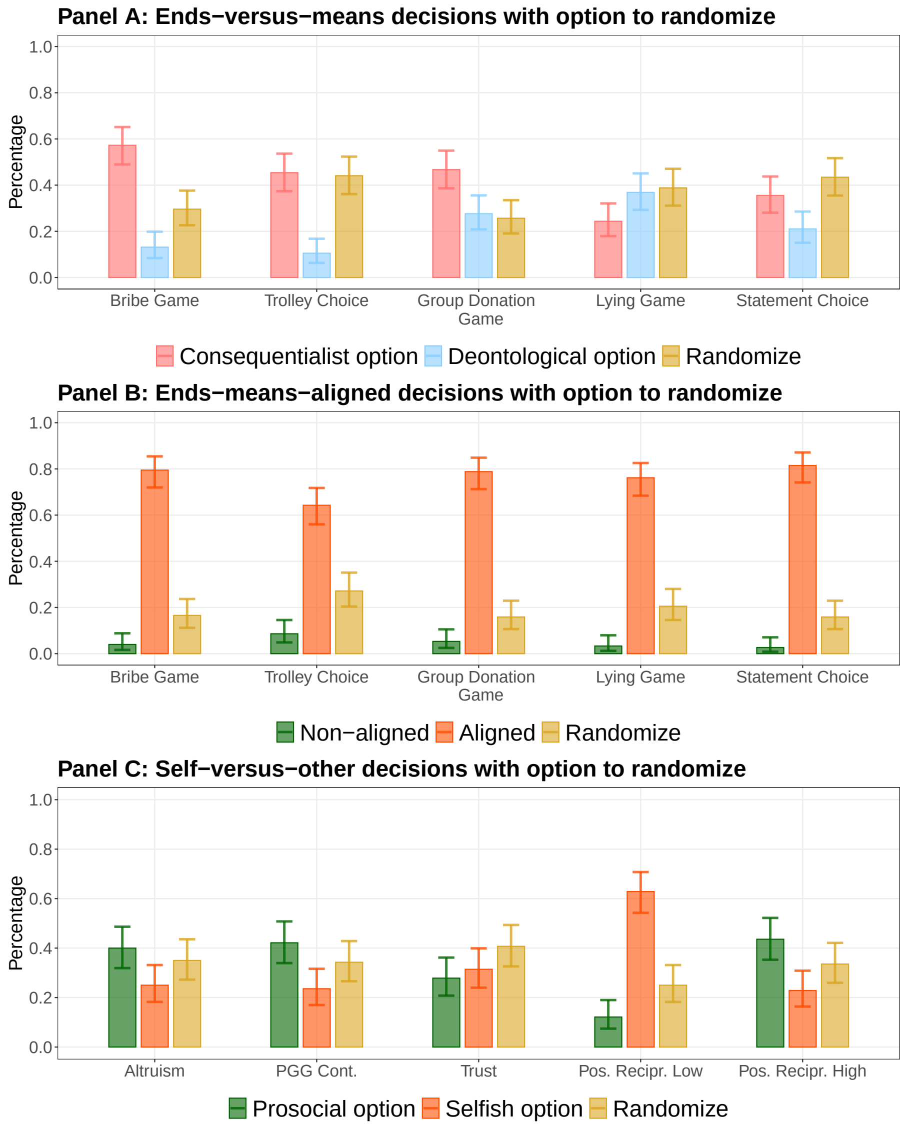
Prevalence (Test 3). We next turn to assessing whether subjects are uncertain of the extent to which they should act as consequentialists or deontologists, and as a result hedge their choices. We test this by analyzing the *extensive* margin of randomization: whether, when allowed to choose a probability of implementing one or the other option, they select some $p \in (0, 1)$. Panel A of Figure 4 shows that preferences for deliberate randomization in the EVM decisions are widespread. In total, 74% of subjects choose to randomize in at least one decision. In each of the five, the fraction randomizing is significantly greater than zero, ranging from 26% in the group donation game to 44% in the trolley dilemma, and averaging to 36% overall. Notably, those who randomize choose a variety of probabilities, with only a minority (around 25% on average across tasks) choosing 50-50. See Appendix Figure H.1 for the distribution of chosen probabilities.

To what extent does the moral conflict inherent in EVM tradeoffs induce deliberate randomization? To answer this question, we compare randomization behavior in EVM tasks to that in their aligned EMA counterparts.[21] On average over the five games, the proportion increases from 19% in EMA to 36% in EVM decisions tasks. (See Appendix Figure H.2 for the distributions of probability choices). The presence of a genuine conflict between ends and means is thus, quite intuitively, a major source of preference uncertainty and hedging. In contrast, *ex-ante* fairness considerations would lead to the same proportions of subjects randomizing in EVM and EMA situations (with flipped probabilities) since the menu of outcomes is the same –only the means of achieving them vary, namely breaking or conforming to a general moral rule.[22]

[21] This difference is statistically significant in each game. The p-values when comparing the randomization rates are $p = 0.01$ for the Bribe Game, $p < 0.01$ for the Trolley dilemma, $p = 0.05$ for the Group Donation Game, $p < 0.01$ for the Lying Game, and $p < 0.01$ for the Statement Choice.

[22] Such fairness concerns may explain why there is still some baseline level of randomization even in EMA choices, particularly for the trolley dilemma. There, subjects may want to give everyone a chance to live.

Figure 4: Distribution of choices in the randomization experiment



Notes: Each bar displays the fraction of subjects choosing the respective option in the described decision-situation. See Sections 3.1.2 and 3.2.2 for details on the situations and the definition of the options. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

Moving to the individual level, we observe consistency on the *extensive* margin of randomization: subjects who randomize in one EVM game are more likely to also randomize in others (as found by Agranov, Healy, and Nielsen (2023) in the different context of decisions under external uncertainty). In Appendix Table H.1, we show that correlations are substantial in all cases, ranging from 0.24 to 0.44, with a Cronbach's alpha of 0.73 (CI=$[0.66, 0.78]$), which is relatively high.

Consistency (Test 4). Given that subjects appear to experience substantial normative uncertainty and ambiguity aversion in EVM choice tasks, we now assess whether the underlying self-knowledge and cautious preferences are stable across situations. As shown in Section 2, we can do so by examining the extent to which randomization probabilities are correlated across tasks. In other words, we now look at consistency on the *intensive* margin of randomization.

Correlations between probabilities are generally small, ranging from -0.03 to 0.25 (see Appendix

Table 4: Correlation matrix for self-versus-other choice tasks

	Altruism Taking	Trust	Recipr. Low	Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.
Altruism	0.65	0.20	0.31	0.38	0.40	0.36	0.23	0.23
	[0.60, 0.70]	[0.12, 0.28]	[0.24, 0.39]	[0.30, 0.45]	[0.33, 0.47]	[0.28, 0.43]	[0.15, 0.31]	[0.15, 0.31]
Altruism Taking		0.16	0.26	0.35	0.36	0.28	0.21	0.21
		[0.08, 0.24]	[0.18, 0.34]	[0.28, 0.42]	[0.28, 0.43]	[0.20, 0.36]	[0.13, 0.29]	[0.13, 0.29]
Trust			0.35	0.45	0.47	0.38	0.30	0.23
			[0.27, 0.42]	[0.38, 0.52]	[0.40, 0.53]	[0.31, 0.45]	[0.22, 0.38]	[0.15, 0.31]
Recipr. Low				0.71	0.36	0.35	0.13	0.14
				[0.66, 0.75]	[0.28, 0.43]	[0.27, 0.42]	[0.05, 0.21]	[0.06, 0.22]
Recipr. High					0.45	0.41	0.30	0.21
					[0.39, 0.52]	[0.34, 0.48]	[0.22, 0.38]	[0.13, 0.29]
PGG Cont.						0.40	0.31	0.22
						[0.33, 0.47]	[0.23, 0.38]	[0.14, 0.30]
Reward Cons.-Int.							0.24	0.55
							[0.16, 0.32]	[0.49, 0.61]
Reward Int.								0.40
								[0.33, 0.47]

Notes: The table displays Pearson correlation coefficients between the self-versus-other choice tasks. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

Table H.2), with a Cronbach's alpha of 0.32 (CI=$[0.16, 0.47]$). Low correlations also persist if we interpret randomization choices as just tendencies, thus coding any probability greater than 50% for one option as a preference for that option (e.g., 80% is coded as 1, and 30% as 0, etc.; we exclude decisions where subjects chose a 50% probability, which as we saw are relatively few. With this binary classification, Cronbach's alpha remain essentially unchanged, equal to 0.30 (CI=$[0.10, 0.49]$). For the correlation matrix, see Appendix Table H.3).

Result 3. *Randomization probabilities in one EVM task are not predictive of those in another. Thus, we find no evidence of stable preferences types along the consequentialist-deontological dimension, even allowing for normative uncertainty and ambiguity aversion.*

5 Prosocial decision-making

5.1 Prevalence and consistency

Prevalence. In our SVO choice tasks, which cover standard games employed by the literature, the results closely match typical patterns found in previous work. The amounts shared in the dictator game, contributed to the public good in the public goods game, sent to the other person in the trust game, and sent back after receiving money are all significantly positive (for every one, $p < 0.001$, one-sample Wilcoxon signed-rank test) and quantitatively similar to previous findings. For instance, in the dictator game, subjects allocate on average 33% of their endowment to the charity, close to the 28% reported by Engel (2011) in his meta-analysis. In the trust game, subjects in the role of the trustor send on average 51% of their endowment, again closely matching the meta-analysis average of 50% reported by Johnson and Mislin (2011). Distributions similarly match standard patterns. Therefore, we confirm the prevalence of both prosocial and selfish decision-making in our sample, in proportions that are in line with the previous literature.

Consistency. To investigate whether these decisions reflect stable preference types along the prosociality-selfishness dimension, we code behaviors so that higher values indicate higher degrees of giving to others, i.e., higher donations in the dictator game, more money sent in the trust game, etc. Table 4 shows that, in stark contrast to EVM situations, behavior in SVO decisions is strongly correlated across games, with most correlations falling between 0.30 and 0.40 and a Cronbach's alpha of $= 0.58\,[0.53, 0.63]$. In particular, the extent of giving to charity in the dictator game consistently predicts the other eight measures. In fact, the lowest correlation of altruism with any other measure is higher than the highest correlation among the EVM tradeoffs. Thus, subjects show a relatively high degree of consistency in the way they resolve self-versus-other tradeoffs.[23]

Result 4. *In line with previous research, we find that both other-regarding and selfish decision-making are prevalent. We further show that behavior in one SVO task is quite predictive of behavior in others, revealing stable preference types along the prosociality-selfishness dimension.*

5.2 Normative uncertainty and cautiousness

Prevalence. To assess the extent to which subjects may be uncertain of how prosocial they are or should be in SVO tradeoffs, we analyze their randomization choices, displayed in Panel C of Figure 4. We observe substantial rates of randomization, similar to those in EVM tradeoffs. This may be somewhat surprising given that latter would seem to involve more difficult judgments, but on the other hand it is consistent with previous findings of high rates of randomization for two SVO tradeoffs. Miao and Zhong (2018) find it in the dictator game and show how it can be explained by ex-ante fairness preferences, while Ong and Qiu (2023) find it by receivers in the ultimatum game and show how it can be explained by max-min preferences with uncertainty over the degree of selfishness or inequity aversion.

Consistency. Turning to the correlation of randomization probabilities across SVO choice tasks, we find high degrees of consistency, as one would expect given Result 4 above. All three cases – (i) extensive margin of randomization, (ii) intensive margin of randomization, (iii) coding randomization choices as tendencies – display relatively high correlation, with Cronbach's alpha values of 0.67 for the extensive margin, 0.67 for the intensive margin, and 0.66 for the case in which we code probabilities as 0 or 1 depending on whether they are either above or below 0.5. For the corresponding correlation tables, see Appendix Tables H.4, H.5, and H.6.

Result 5. *Subjects randomize their choice in SVO situations to about the same degree as in EVM settings, revealing again substantial normative uncertainty and cautiousness. However, randomization probabilities in one SVO task are quite predictive of those in another, in contrast to EVM. There are thus*

[23]For altruism (dictator game), trust (first-mover trust game), and both reciprocity measures (second-mover trust game), we can compare our individual-level correlations with those in Chapman et al. (2023). They report for altruism (measurement-error corrected) correlations of 0.34 with both reciprocity measures and 0.60 with trust. Their correlation coefficient of trust with both reciprocity measures is 0.49, and the correlation between low and high reciprocity is 0.86. Hence, with the exception of the relation of altruism with trust, our correlations are very similar to theirs, even though they use a representative sample of the US population, while our sample contains mostly German students.

stable (cautious-) preferences types in terms of prosociality/selfishness, but not in terms of deontologism/consequentialism.

6 Comparing ends-versus-means and self-versus-other decision-making

We now turn to a direct comparison of behaviors in EVM and SVO tradeoffs. While our primary focus is on the novel EVM dilemmas, the comparison serves to better contextualize the degree of consistency observed within each domain.

6.1 Comparing consistency

Comparing correlations. As seen previously, correlations between tasks reveal substantial consistency in SVO decisions and almost none in EVM ones. In the aggregate, not only is Cronbach's alpha high in one case and low in the other, but the confidence intervals (CI) do not even overlap: for EVM, alpha $= 0.20$ with CI = $[0.09, 0.30]$, whereas for SVO alpha $= 0.58$ with CI = $[0.53, 0.63]$

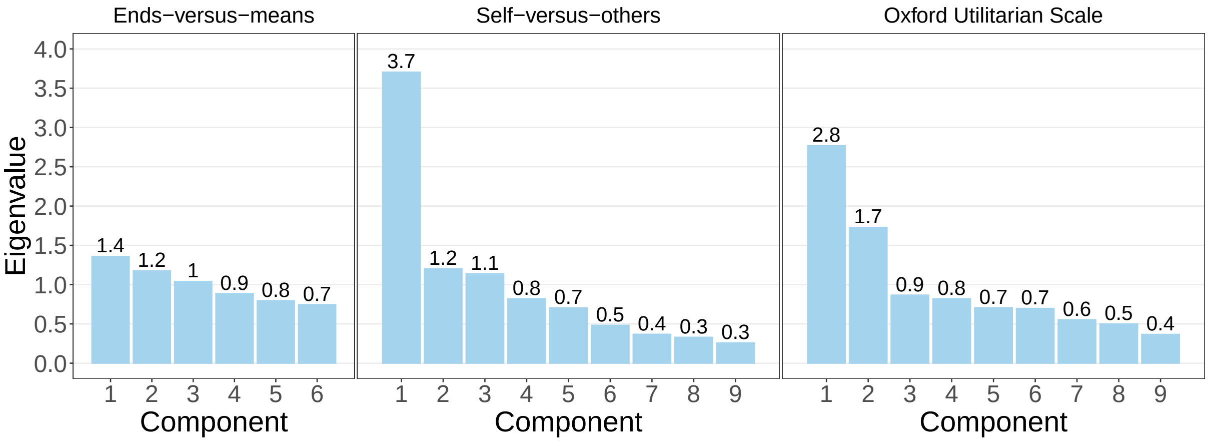
Comparing principal-component analyses. As a second method of comparison, we use principal component analysis (PCA) to examine how well behavior in EVM and SVO can be explained by a smaller subset of variables. As a reference, we also perform a PCA on the Oxford Utilitarian scale.

Figure 5 displays, for each block, the eigenvalues of the corresponding components. For the EVM block they are fairly close to each other, whereas for the SVO block the first component has a substantially higher eigenvalue than the others. Thus, one component, which might be termed "general prosociality" explains subjects' SVO choices rather well, whereas no such dominant component exists for the EVM choices, confirming that these are largely independent.[24] The first principal component for the SVO block is even stronger than that for the OUS questionnaire block, which is quite notable since the questions in the latter are largely rephrasings of the same one, namely whether it is permissible to harm one person for some greater good. This last fact is also why the OUS measure is strongly correlated with the trolley choices, as seen previously, and not correlated with any of the other EVM decisions (see Appendix Table G.2 for details).

Robustness. The variables in the EVM block are dichotomous, while those in the SVO block are more fine-grained. Accordingly, relying on standard Pearson correlation coefficients could potentially confound the comparison of the two blocks. To alleviate this concern, we replicate our results with dichotomized SVO variables, see Appendix Section G.1 for details.

[24]This result is not specific to the fact that we have six EVM situations and nine SVO measures. We also conducted a PCA for every combination of six situations out of the nine SVO measures: in every one of the resulting 84 combinations, the variance explained by the first component of these six situations is twice as high as the first component can explain in the six EVM situations. The resulting eigenvalues from this exercise range from 2.36 to 3.02. Hence, no matter which SVO situations are taken, a dimensionality reduction always performs better among SVO tradeoffs than among EVM situations, respectively.

Figure 5: Principal component analyses



Notes: The figure displays the results of principal-component analyses conducted on the six ends-versus-means and nine self-versus-other choice tasks, as well as the nine items of the Oxford Utilitarian scale. The y-axis shows the eigenvalues of each component. Higher eigenvalues indicate that the respective component explains more variance in the data.

6.2 Are ends-versus-means and self-versus other behaviors related?

Lastly, we examine whether behavior in SVO choice situations is distinct from, or systematically related to, that in EVM dilemmas. Given that we found no consistency across EVM tasks, we would not expect any general relationship, but it could still be the case that some EVM choices are predictive of some SVO ones.

Starting with the trolley dilemma given its importance in the literature, we find no relationship with any of the SVO choices. Between subjects choosing the consequentialist and those choosing the deontological option, average giving to the charity in the dictator game differs by just 0.13 Euros ($p = 0.86$, two-sample t-test), average amounts sent in the trust game by 0.10 Euros ($p = 0.61$), and contributions to the public good by 0.02 Euros ($p = 0.91$).[25]

In Table 5, we analyze the relationship more generally by presenting the pairwise correlations between all EVM and all SVO games. We find that the two blocks are largely unrelated: all pairwise correlations are below 0.20, with the large majority being close to zero.[26]

Result 6. *The level of consistency across SVO decisions is substantially higher than that across EVM decisions. Moreover, the two types of moral decision-making are unrelated: none of the EVM choices predicts any SVO choice, and vice versa.*

7 Conclusion and directions for further research

Our results bring substantial nuance to the classical dichotomy opposing Consequentialism and Deontologism. On one hand, deontological decision-making is a robust phenomenon, not an artifact

[25]Similarly, Capraro et al. (2018) find no differences in dictator and trust game behavior between subjects choosing the consequentialist and deontological option in the trapdoor version of the trolley dilemma.

[26]We replicate the finding of low correlations also when using point-biserial correlations, which are designed for pairwise comparisons between dichotomous and continuous variables, see Appendix Table G.7.

Table 5: Correlations between ends-versus-means and self-versus-other choice tasks

	Altruism	Altruism Taking	Trust	Recipr. Low	Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.
SAL Trolley	-0.01 [-0.09, 0.08]	-0.03 [-0.11, 0.06]	0.02 [-0.06, 0.11]	-0.05 [-0.13, 0.04]	-0.03 [-0.12, 0.05]	0.00 [-0.08, 0.09]	-0.02 [-0.10, 0.06]	0.05 [-0.03, 0.14]	-0.02 [-0.10, 0.06]
Group. Don.	0.05 [-0.04, 0.13]	0.03 [-0.06, 0.11]	-0.04 [-0.13, 0.04]	0.14 [0.05, 0.22]	0.05 [-0.04, 0.13]	-0.02 [-0.10, 0.06]	0.09 [0.00, 0.17]	-0.03 [-0.12, 0.05]	-0.02 [-0.10, 0.07]
Bribe	-0.04 [-0.12, 0.05]	-0.02 [-0.11, 0.06]	-0.11 [-0.20, -0.03]	-0.06 [-0.14, 0.03]	-0.15 [-0.23, -0.07]	-0.06 [-0.14, 0.02]	-0.06 [-0.14, 0.03]	-0.18 [-0.26, -0.10]	-0.18 [-0.26, -0.10]
Statement Choice	0.00 [-0.08, 0.09]	0.01 [-0.08, 0.09]	-0.12 [-0.20, -0.03]	0.04 [-0.05, 0.12]	-0.07 [-0.15, 0.01]	-0.00 [-0.09, 0.08]	0.05 [-0.03, 0.14]	-0.15 [-0.23, -0.07]	-0.04 [-0.12, 0.05]
Lying	-0.03 [-0.11, 0.06]	-0.04 [-0.12, 0.04]	-0.03 [-0.11, 0.06]	-0.03 [-0.11, 0.06]	0.00 [-0.08, 0.08]	-0.05 [-0.13, 0.04]	-0.00 [-0.09, 0.08]	-0.02 [-0.11, 0.06]	-0.01 [-0.09, 0.08]
Rule following	0.11 [0.03, 0.20]	0.06 [-0.02, 0.14]	0.03 [-0.05, 0.11]	0.11 [0.02, 0.19]	0.10 [0.02, 0.19]	0.03 [-0.05, 0.12]	0.11 [0.03, 0.20]	-0.01 [-0.09, 0.08]	0.13 [0.04, 0.21]

Notes: The table displays Pearson correlation coefficients between the ends-versus-means and self-versus-other choice tasks. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

of hypothetical choices or unrealistic dilemmas: we find it in a real-stakes version of the trolley and in five other games pitting socially desirable ends against arguably (and, for our subjects, *de facto*) objectionable means. Treating all economic agents as solely consequentialist thus appears to be an inaccurate representation of their preferences and decisions.

On the other hand, in our subject population there appear to be no consistently deontological individuals and only a small fraction of consistently consequentialist ones –more generally, no consistent ordering of participants into more or less deontological types. Instead, most participants' choices across the six ends-versus-means dilemmas are largely uncorrelated. Subjects clearly feel the pull of both moral principles, resulting in significant normative uncertainty as revealed by their desire to randomize, but they appear to resolve it in a very situation-specific manner. These behaviors reveal that they disagree on how important they consider each moral rule to be, even fixing consequences as in our four core choice tasks.

The high degree of context-dependency across ends-versus-means choices means that behavior in one setting cannot readily be used to predict behavior in another. As we show, this holds particularly for the workhorse dilemma used in the literature to understand moral behavior – the trolley dilemma. These findings stand in sharp contrast to studies such as Helzer et al. (2017) and Rehren and Sinnott-Armstrong (2023) that focus only on (hypothetical) sacrificial dilemmas and find generally high correlations across time and framings.[27] Our results show that robustness to different framings of trolley-like dilemmas, which we also find, does not allow extrapolating conclusions to other settings where consequentialism and deontologism conflict.

In contrast, inconsistency is much less of an issue in self-versus-other tradeoffs. Here, the workhorse game used in the literature – the dictator game – predicts behavior in other self-versus-other settings

[27]The first paper asked subjects twice for their judgments of moral appropriateness in six scenarios involving the same tradeoff of sacrificing one life to save five; it found significant correlations both over time and between scenarios. The second paper asked similar "1 life for 5" sacrificial-dilemmas questions thrice and found relatively high temporal correlations for each question, but with substantial heterogeneity: for some questions, subjects' judgments were very stable, for others they showed substantial shifts or even reversals, sometimes twice in a row. Relatedly, Graham et al. (2013) found high test-retest reliability for answers to the Moral Foundations questionnaire.

well. As such, extrapolation to other settings or choice tasks appears to be possible. For instance, experimentally measured altruism predicts prosocial behavior in the field (e.g., Rustagi, Engel, and Kosfeld, 2010; Falk et al., 2018; Kosse and Tincani, 2020; Chen et al., 2021; Lades, Laffan, and Weber, 2021).

A first direction for further research, clearly suggested by the contrast drawn above with findings within the specific realm of sacrificial dilemmas, would be to complement our study of consistency across EVM situations by examining the stability of choices, both deterministic and randomized, between repetitions of the same task over time. Knowing which of the six incentivized behaviors we study display temporal consistency and which do not would shed further light on the existence of stable preference types and the ability of our games to detect them.

A second important direction is the role of image. As mentioned in the introduction, there is some evidence that people who make deontological decisions are regarded by others as more moral and trustworthy than those who act consequentially. Our results on the lack of correlation across the EVM and SVO domains indicate that this differential esteem is likely unwarranted. But if people nonetheless have such beliefs, this creates a reputational incentive to be seen as acting deontologically. This mechanism could be tested by examining whether decisions in EVM situations become more deontological when choices are observed or even publicized than when they are anonymous.

On the methodological side, the paper introduces a new experimental paradigm – ends-versus-means games, as fundamentally distinct from the self-versus-other games used to measure prosociality. It is quite flexible, and in particular independent of trolley-like sacrificial dilemmas, a feature which turns out to be important. While another contribution of the paper is to make trolley choices "real," this particular task could be dropped without changing the general architecture or the overall results of our study. Similarly, a researcher could replace one or more of our tests of Consequentialism vs. Deontologism (lying game, bribing game, non-pivotal group choice, aversive statement, rule following) with others they deemed preferable, or add new ones. The methodology of examining the consistency of decisions within that ends-versus-means block, and potential correlations with behavior in a block of standard self-versus-others prosociality games, would remain the same.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019). "Preferences for Truth-Telling". *Econometrica* 87 (4): 1115–53. [5]

Agranov, Marina, Paul J. Healy, and Kirby Nielsen (2023). "Stable Randomisation". *Economic Journal* 133 (655): 2553–79. [4, 24]

Agranov, Marina, and Pietro Ortoleva (2017). "Stochastic Choice and Preferences for Randomization". *Journal of Political Economy* 125 (1): 40–68. [14]

Agranov, Marina, and Pietro Ortoleva (2022). "Revealed Preferences for Randomization: An Overview". *AEA Papers and Proceedings* 112: 426–30. [6, 8]

Agranov, Marina, and Pietro Ortoleva (2023). "Ranges of Randomization". *Review of Economics and Statistics*, (7): 1–44. [3, 14]

Alexander, Larry, and Michael Moore (2016). *Deontological Ethics*. Edward N. Zalta, ed. The Stanford Encyclopedia of Philosophy. [1]

Alger, Ingela, and Jörgen W. Weibull (2013). "Homo Moralis–Preference Evolution Under Incomplete Information and Assortative Matching". *Econometrica* 81 (6): 2269–302. [5]

Alger, Ingela, and Jörgen W. Weibull (2016). "Evolution and Kantian Morality". *Games and Economic Behavior* 98 (7): 56–67. [5]

Alger, Ingela, Jörgen W. Weibull, and Laurent Lehmann (2020). "Evolution of Preferences in Structured Populations: Genes, Guns, and Culture". *Journal of Economic Theory* 185: 1–52. [5]

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean François Bonnefon, and Iyad Rahwan (2018). "The Moral Machine Experiment". *Nature* 563 (7729): 59–64. [2]

Awad, Edmond, Sohan Dsouza, Azim Shariff, Iyad Rahwan, and Jean François Bonnefon (2020). "Universals and Variations in Moral Decisions Made in 42 Countries by 70,000 Participants". *Proceedings of the National Academy of Sciences of the United States of America* 117 (5): 2332–37. [6, 18]

Bartels, Daniel M. (2008). "Principled Moral Sentiment and the Flexibility of Moral Judgment and Decision Making". *Cognition* 108 (2): 381–417. [6]

Bauman, Christopher W., A. Peter Mcgraw, Daniel M. Bartels, and Caleb Warren (2014). "Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology". *Social and Personality Psychology Compass* 8 (9): 536–54. [6, 11]

Bénabou, Roland, Armin Falk, Luca Henkel, and Jean Tirole (2023). "Eliciting Moral Preferences under Image Concerns: Theory and Experiment". *Working Paper*, [5, 13, 20, 48]

Bénabou, Roland, and Jean Tirole (2011). "Identity, Morals, and Taboos: Beliefs as Assets". *Quarterly Journal of Economics* 126 (2): 805–55. [5, 13]

Bentham, Jeremy (1789). *An Introduction to the Principles of Morals*. London: Athlone. [1]

Biziou-van-Pol, Laura, Jana Haenen, Arianna Novaro, Andrés Occhipinti Liberman, and Valerio Capraro (2015). "Does Telling White Lies Signal Pro-Social Preferences?" *Judgment and Decision Making* 10 (6): 538–48. [5]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "Hroot: Hamburg Registration and Organization Online Tool". *European Economic Review* 71: 117–20. [16]

Bostyn, Dries H., Sybren Sevenhant, and Arne Roets (2018). "Of Mice, Men, and Trolleys: Hypothetical Judgment Versus Real-Life Behavior in Trolley-Style Moral Dilemmas". *Psychological Science* 29 (7): 1084–93. [6]

Bursztyn, Leonardo, Bruno Ferman, Ali Hasanain, Michael Callen, Saad Gulzar, and Noam Yuchtman (2020). "Political Identity: Experimental Evidence on Anti-Americanism in Pakistan". *Journal of the European Economic Association* 18 (5): 2532–60. [13, 42]

Cappelen, Alexander W., Erik Ø. Sørensen, and Bertil Tungodden (2013). "When Do We Lie?" *Journal of Economic Behavior & Organization* 93 (9): 258–65. [5]

Capraro, Valerio, Jonathan Sippel, Bonan Zhao, Levin Hornischer, Morgan Savary, Zoi Terzopoulou, Pierre Faucher, and Simone F. Griffioen (2018). "People Making Deontological Judgments in the Trapdoor Dilemma Are Perceived to Be More Prosocial in Economic Games than They Actually Are". *PLOS ONE* 13 (10): Pablo Brañas-Garza, ed., e0205066. [4, 28]

Cerreia-Vioglio, Simone, David Dillenberger, and Pietro Ortoleva (2024). "Caution and Reference Effects". *Econometrica* 92 (6): 2069–103. [6, 7]

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer (2023). "Econographics". *Journal of Political Economy Microeconomics* 1 (1): 115–61. [26]

Chen, Daniel L., and Martin Schonger (2022). "Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations". *Science Advances* 8 (19): eabb3925. [5]

Chen, Daniel L., Martin Schonger, and Chris Wickens (2016). "oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments". *Journal of Behavioral and Experimental Finance* 9: 88–97. [16]

Chen, Zihua, Ximeng Fang, Timo Freyer, Lorenz Goette, and Chui Yee Ho (2021). "Prosociality Predicts Individual Behavior and Collective Outcomes in the COVID-19 Pandemic". *Working Paper*, [30]

Costa, Albert, Alice Foucart, Sayuri Hayakawa, Melina Aparici, Jose Apesteguia, Joy Heafner, and Boaz Keysar (2014). "Your Morals Depend on Language". *PLoS ONE* 9 (4): [6]

Dietrich, Franz, and Brian Jabarian (2022). "Decision under normative uncertainty". *Economics and Philosophy* 38 (3): 372–94. [2]

Engel, Christoph (2011). "Dictator Games: A Meta Study". *Experimental Economics* 14 (4): 583–610. [25]

Enke, Benjamin, and Thomas Graeber (2023). "Cognitive Uncertainty". *Quarterly Journal of Economics* 138 (4): 2021–67. [2]

Enke, Benjamin, Thomas Graeber, and Ryan Oprea (2023). "Complexity and Time". *Working Paper*, [2]

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann (2022). "Moral Universalism: Measurement and Economic Relevance". *Management Science* 68 (5): 3590–603. [3, 16]

Erat, Sanjiv, and Uri Gneezy (2012). "White Lies". *Management Science* 58 (4): 723–33. [5, 39]

Everett, Jim A.C., Nadira S. Faber, Julian Savulescu, and Molly J. Crockett (2018). "The Costs of Being Consequentialist: Social Inference from Instrumental Harm and Impartial Beneficence". *Journal of Experimental Social Psychology* 79 (August): 200–16. [4]

Everett, Jim A.C., David A. Pizarro, and Molly J. Crockett (2016). "Inference of Trustworthiness From Intuitive Moral Judgments". *Journal of Experimental Psychology: General* 145 (6): 772–87. [4]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2018). "Global Evidence on Economic Preferences". *Quarterly Journal of Economics* 133 (4): 1645–92. [30]

Falk, Armin, Ernst Fehr, and Urs Fischbacher (2003). “On the Nature of Fair Behavior”. *Economic Inquiry* 41 (1): 20–26. [5]

Falk, Armin, Ernst Fehr, and Urs Fischbacher (2008). “Testing Theories of Fairness-Intentions Matter”. *Games and Economic Behavior* 62 (1): 287–303. [5]

Falk, Armin, and Urs Fischbacher (2006). “A Theory of Reciprocity”. *Games and Economic Behavior* 54 (2): 293–315. [5]

Falk, Armin, and Thomas Graeber (2020). “Delayed Negative Effects of Prosocial Spending on Happiness”. *Proceedings of the National Academy of Sciences* 117 (12): 6463–68. [11, 36]

Falk, Armin, Thomas Neuber, and Nora Szech (2020). “Diffusion of Being Pivotal and Immoral Outcomes”. *Review of Economic Studies* 87 (5): 2205–29. [12, 40, 41]

Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni (2009). “Moral Bias in Large Elections: Theory and Experimental Evidence”. *American Political Science Review* 103 (2): 175–92. [5]

Feddersen, Timothy, and Alvaro Sandroni (2006). “A Theory of Participation in Elections”. *American Economic Review* 96 (4): 1271–82. [5]

Feess, Eberhard, Florian Kerzenmacher, and Yuriy Timofeyev (2022). “Utilitarian or Deontological Models of Moral Behavior - What Predicts Morally Questionable Decisions?” *European Economic Review* 149: 104264. [5]

Foot, Philippa (1967). “The Problem of Abortion and the Doctrine of the Double Effect”. *Oxford Review* 5: 5–15. [2]

Gächter, Simon, Lucas Molleman, and Daniele Nosenzo (1, 2025). “Why People Follow Rules”. *Nature Human Behaviour* 9 (7): 1342–54. [5]

Gawronski, Bertram, and Jennifer S. Beer (2017). “What Makes Moral Dilemma Judgments ”utilitarian” or ”deontological”?” *Social Neuroscience* 12 (6): 626–32. [6]

Gibson, Rajna, Carmen Tanner, and Alexander F Wagner (2013). “Preferences for Truthfulness: Heterogeneity among and within Individuals”. *American Economic Review* 103 (1): 532–48. [5]

Gneezy, Uri (2005). “Deception: The Role of Consequences”. *American Economic Review* 95 (1): 384–94. [38]

Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel Van Veldhuizen (2020). “Bribing the Self”. *Games and Economic Behavior* 120 (3): 311–24. [12]

Gold, Natalie, Briony D. Pulford, and Andrew M. Colman (2015). “Do as I Say, Don’t Do as I Do: Differences in Moral Judgments Do Not Translate into Differences in Decisions in Real-Life Trolley Problems”. *Journal of Economic Psychology* 47: 50–61. [6]

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto (2013). “Moral Foundations Theory”. In *Advances in Experimental Social Psychology*. vol. 47, Elsevier, 55–130. [29]

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): 366–85. [3, 16]

Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen (2001). “An fMRI Investigation of Emotional Engagement in Moral Judgment”. *Science* 293 (5537): 2105–8. [6]

Hauser, Marc, Fiery Cushman, Liane Young, R. Kang-Xing Jin, and John Mikhail (2007). "A Dissociation between Moral Judgments and Justifications". *Mind and Language* 22 (1): 1–21. [6]

Helzer, Erik G., William Fleeson, R. Michael Furr, and Maxwell Barranti (2017). "Once a Utilitarian, Consistently a Utilitarian? Examining Principledness in Moral Judgment via the Robustness of Individual Differences". *Journal of Personality* 85 (4): 505–17. [29]

Hsu, Ming, Cédric Anen, and Steven R. Quartz (2008). "The Right and the Good: Distributive Justice and Neural Encoding of Equity and Efficiency". *Science* 320 (5879): 1092–95. [6]

Johnson, Noel D., and Alexandra A. Mislin (2011). "Trust Games: A Meta-Analysis". *Journal of Economic Psychology* 32 (5): 865–89. [25]

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): 131–64. [3, 15, 22]

Kant, Immanuel (1785). *Grundlegung Zur Metaphysik Der Sitten*. J. F. Hartknoch. [1, 7, 12]

Kant, Immanuel (1788). "Über den Gebrauch teleologischer Prinzipien in der Philosophie". In *Der Streit der Fakultäten*. vol. 8, Kants gesammelte Schriften Berlin: Preußische Akademie der Wissenschaften. Akademie-Ausgabe (AA) 8:420. [11]

Kant, Immanuel (1797a). *Die Metaphysik der Sitten*. vol. 6, Kants gesammelte Schriften. Berlin: Preußische Akademie der Wissenschaften. Akademie-Ausgabe (AA) 6:450. [12]

Kant, Immanuel (1797b). "Über ein vermeintes Recht aus Menschenliebe zu lügen". In *Der Streit der Fakultäten*. vol. 8, Kants gesammelte Schriften Berlin: Preußische Akademie der Wissenschaften. Akademie-Ausgabe (AA) 8:427. [12]

Kimbrough, Erik O., and Alexander Vostroknutov (2016). "Norms Make Preferences Social". *Journal of the European Economic Association* 14 (3): 608–38. [5, 13, 42]

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005). "A Smooth Model of Decision Making under Ambiguity". *Econometrica* 73 (6): 1849–92. [7]

Kosse, Fabian, and Michela M. Tincani (2020). "Prosociality Predicts Labor Market Success around the World". *Nature Communications* 11 (1): 5298. [30]

Kranz, Sebastian (2010). "Moral Norms in a Partly Compliant Society". *Games and Economic Behavior* 68 (1): 255–74. [5]

Lades, Leonhard K., Kate Laffan, and Till O. Weber (2021). "Do Economic Preferences Predict Pro-Environmental Behaviour?" *Ecological Economics* 183 (5): 106977. [30]

Laffont, Jean-Jacques (1975). "Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics". *Economica* 42 (168): 430. [5]

Lanteri, Alessandro, Chiara Chelini, and Salvatore Rizzello (2008). "An Experimental Investigation of Emotions and Reasoning in the Trolley Problem". *Journal of Business Ethics* 83 (4): 789–804. [6]

Maccheroni, Fabio (2002). "Maxmin under Risk". *Economic Theory* 19 (3): 623–43. [7]

Miao, Bin, and Songfa Zhong (2018). "Probabilistic social preference: how Machina's Mom randomizes her choice". *Economic Theory* 65 (1): 1–24. [26]

Milgrom, Paul R., and Robert J. Weber (1982). "A Theory of Auctions and Competitive Bidding". *Econometrica* 50 (5): 1089–122. [9]

Mill, John Stuart (1863). *Utilitarianism*. Son & Bourn Parker, ed. London: West Strand. [1]

Ong, Qiyan, and Jianying Qiu (2023). “Paying for randomization and indecisiveness”. *Journal of Risk and Uncertainty* 67 (1): 45–72. [2, 26]

Rai, Tage S., and Keith J. Holyoak (2010). “Moral Principles or Consumer Preferences? Alternative Framings of the Trolley Problem”. *Cognitive Science* 34 (2): 311–21. [6]

Rehren, Patrick, and Walter Sinnott-Armstrong (2023). “How Stable Are Moral Judgments?” *Review of Philosophy and Psychology* 14: 1377–403. [29]

Roemer, John E. (2010). “Kantian Equilibrium”. *Scandinavian Journal of Economics* 112 (1): 1–24. [5]

Rustagi, Devesh, Stefanie Engel, and Michael Kosfeld (2010). “Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management”. *Science* 330 (6006): 961–65. [30]

Sinnott-Armstrong, Walter (2019). *Consequentialism*. Edward N. Zalta, ed. The Stanford Encyclopedia of Philosophy. [1]

Thompson, Judith Jarvis (1976). “Killing, Letting Die, and the Trolley Problem”. *Monist* 59 (2): 204–17. [2]

Van Leeuwen, Boris, and Ingela Alger (2021). “Estimating Social Preferences and Kantian Morality in Strategic Interactions”. *Working Paper*, [5]

ONLINE APPENDIX

A Formalizing self-versus-other tradeoffs and preferences

Here, we provide the details of the extension of our framework to self-versus-other (SVO) tradeoffs.

Choices. Whereas previously in Section 2.1, actions solely affected others, we now introduce personal consequences. In task t, an agent chooses between a selfish action and a prosocial one with respective payoffs to themselves and others equal for simplicity to $(x_t, 0)$ if $a = s$ and $(0, y_t)$ if $a = p$. Dictator games and charitable donations, for instance, fit this setting. We let $\pi(a)$ denote an indicator equal to 1 if $a = p$.

Preferences. We start as before with pure types, either selfish or altruistic. When selfish, the agent values his own payoff as $\Phi_t^i x_t$ and others' payoff as $\phi_t^i y_t < \Phi_t^i x_t$; (ii) when prosocial, these valuations are instead $\phi_t^i x_t$ and $\Phi_t^i y_t > \phi_t^i x_t$, and moreover the agent experiences negative utility (e.g., guilt) $-\omega_t^i y_t$ when his choice is selfish and thereby deprives others of y_t. Accordingly, selfish and prosocial preferences take the following form:

$$U_{s,t}^i(a) = \pi(a)\phi_t^i y_t + (1 - \chi(a))\Phi_t^i x_t,$$
$$U_{p,t}^i(a) = \pi(a)\Phi_t^i y_t + (1 - \pi(a))(\phi_t^i x_t - \omega_t^i y_t).$$

Mixed types then have utility $U_t^i(a) = \alpha^i U_{s,t}^i(a) + (1 - \alpha^i)U_{p,t}^i(a)$ and those with unknown preferences are either selfish or prosocial with probabilities $(\alpha^i, 1 - \alpha^i)$ and ambiguity aversion η^i. All the previous steps of Section 2 go through, simply substituting $(\Phi_t^i x_t, \phi_t^i y_t, \Phi_t^i y_t, \omega_t^i - \Phi_t^i x_t)$ for $(V^i, t, v_t^i, k_t^i, K_t^i)$. Thus, agents with mixed preference will behave altruistically in task t if their $\alpha^i/(1 - \alpha^i)$ exceeds some threshold $1/z_t^i$ given by (2) and those with unknown preferences will randomize with probability q_t^i given by (4), with the appropriate substitutions.

B Details on ends-versus-means choice tasks

In this section, we provide the full description of the ends versus means choice tasks covered in Section 3.1.

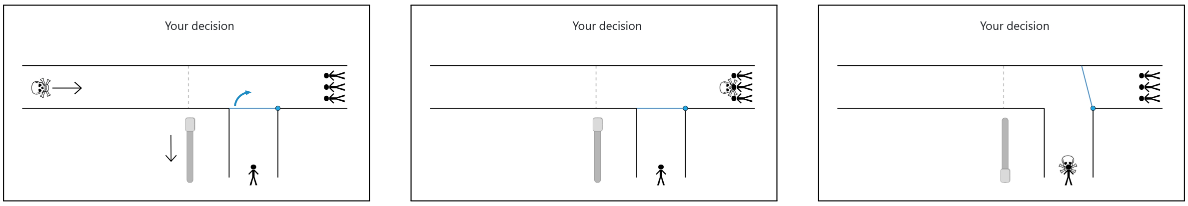
B.1 Trolley dilemma

We implement the trolley dilemma with real, meaningful consequences using the *Saving a Life* Paradigm (Falk and Graeber, 2020). Key features of our version are that choices involve the saving of actual human lives, and that they occur in a realistic context that nonetheless contains all relevant features of the trolley dilemma.

Saving a Life (SAL) paradigm. We partnered with the non-profit organization *Operation ASHA*, which treats people in India suffering from tuberculosis using innovative methods and procedures. Tuberculosis is a highly lethal infectious disease if untreated, but curable with a high success rate if

Figure B.1: Decision screen of the trolley dilemma

(a) Initial position | (b) End position not pulling the slider | (c) End position pulling the slider



Notes: The figure displays screenshots of the animation used for the trolley dilemma decision. Panel a) shows the starting position from which the skull symbol slowly moves to the right. The arrows are included for illustration and were not part of the experimental decision screen. Panel b) shows the end position if a subject decides not to pull the slider, indicating no redirection of the donation. Panel c) shows the end position if a subject decides to pull the slider, indicating a redirection of the donation.

treated. An amount of 380 Euros allows the charity to treat five people suffering from tuberculosis, one of which on average would have otherwise died. Thus, each donation implemented saves one life in expectation. This calculation takes into account treatment success rates, other fatality rates, and alternative treatment possibilities, using peer-reviewed epidemiological studies and information about the cost structure of the charity. Subjects receive detailed information on the context of the SAL choice paradigm and all aspects of the computation.

Based on this setup, we implement a dilemma very similar to the classical Trolley. Prior to the experiment, the charity identified people suffering from tuberculosis who could be treated in two distinct Indian states, A and B.[28] The situation described to subjects was then the following.

- A donation of 380 Euros has been *preset* to cover 5 people suffering from tuberculosis in State A. If it is actually implemented (after the experiment), none of the five will die from the disease. If not implemented, one will die, in expectation. Donation to State A is thus the default.

- In State B, 15 people are suffering from tuberculosis, for whom no donation is initiated. If the donation preset for State A is redirected to State B, the amount will be multiplied by three. With those 1,140 Euros, 15 people will be treated, thus saving (in expectation) three lives that would otherwise be lost to tuberculosis, but foregoing the saving of one in State A.

Decision. Subjects have the choice of whether to redirect the donation from State A to State B or not. They can do so within an animation, a screenshot of which is depicted in Figure B.1 (arrows added for exposition). Absent any action, the skull symbol proceeds from left to right along the track and will eventually hit the three figures representing lives in State B. The subject can, however, redirect the skull toward the one figure representing a life in State A, by pulling a lever that will cause a gate (drawn in blue in Panel A) to pivot, provided this is done before the skull has passed the gate. Subjects thus have two options:

- Not pulling the slider: no redirection of donation $\rightarrow$ three lives in State B are lost to tuberculosis.

- Pulling the slider: redirecting donation $\rightarrow$ one life in State A is lost to tuberculosis.

[28]We used the states of Maharashtra and Orissa, in which the charity operates. We randomized between subjects which state was State A and which State B.

Tradeoff. The moral dilemma involved is very similar to the classical trolley dilemma. According to deontological ethics, it is not permissible to redirect the donation (or, equivalently, the skull), thereby causing the death of (more precisely, not saving) one person in State A in order to save three in State B. This represents active harm to a human being, and arguably using them as a means to an end. According to consequentialist cost-benefit analysis, on the other hand, saving three rather than one is the right thing to do, and no different from tradeoffs implicitly made every day in medicine, product or road safety, and public-budget allocations.

Treatments. Our SAL version of the trolley dilemma differs from moral philosophers' classical one in two ways. First, the scenario is not about imagined train tracks and repairmen happening to be on them, but about actual patients and treatments in the context of a real and common disease. To investigate the effect of this added realism, subjects also faced the classical train-track trolley version at some other stage in the experiment (a week apart, and in randomized order). Second, and most importantly, subjects' choices are not hypothetical but can have real, important consequences. To properly isolate the effect of this latter feature, we ran two different between-subject treatments. In Treatment *SAL-Hypothetical*, the choice environment is exactly as described above, but all choices are, and are presented as, hypothetical: there is never any actual donation. In contrast, in Treatment *SAL-Real*, for each subject, there is a 10% probability that their decision will be implemented, resulting in either one life saved in State A or three in State B.[29]

***Aligned* version.** A donation sufficient to treat 15 people and thus save three lives in State A is preset, and none in State B. If the donation is redirected, it is reduced so that only five people will be treated in State B, thus saving one life there but sacrificing three in State A. Consequentialist cost-benefit analysis and the deontological directive against using a human being as a means to an end now both agree to not redirect the donation, thus prescribing to not pulling the slider.

B.2 Lying game

This game was adapted from the classical sender-receiver game in Gneezy (2005), which we modify from a self-versus-other to an ends-versus-means dilemma. The receiver must choose between options A and B without knowing anything about the consequences of either option. The only piece of information they receive is a message from the sender, who knows that if the receiver chooses option A, 15 Euros will be donated to the children's cancer charity, whereas choosing B will earn the receiver 2 Euros. As with all the EVM games, donations and transfers are paid by the experimenter, eliminating any self-versus-other tradeoff for the decision-maker. The sender can send either of the following messages:

- Message 1: "Option A will give you the higher personal payment" (lie).
- Message 2: "Option B will give you the higher personal payment" (truth).

[29] We ended up implementing the decisions of 23 subjects. As a result of the ensuing donations in both states, 265 patients were treated, and thus 53 lives were saved, in expectation.

The sender is informed that in more than 90% of cases, receivers choose the option mentioned in the message,[30] and that they will never know whether the message was true or false, nor what situation the sender faced. The outcome of interest concerns the decisions of senders, whereas the behavior of subjects playing the role of receivers is not part of the analysis. Hence, subjects of the main experiment take the sender's role, while receivers are part of a separate sample.

Tradeoff. The game puts subjects in a situation in which they need to lie (to someone else's minor detriment) in order to trigger the more socially valuable donation. Such "white lies" are justifiable by consequentialist principles but not under deontological ethics (Erat and Gneezy, 2012). Accordingly, the former prescribes Message 1, the latter Message 2.

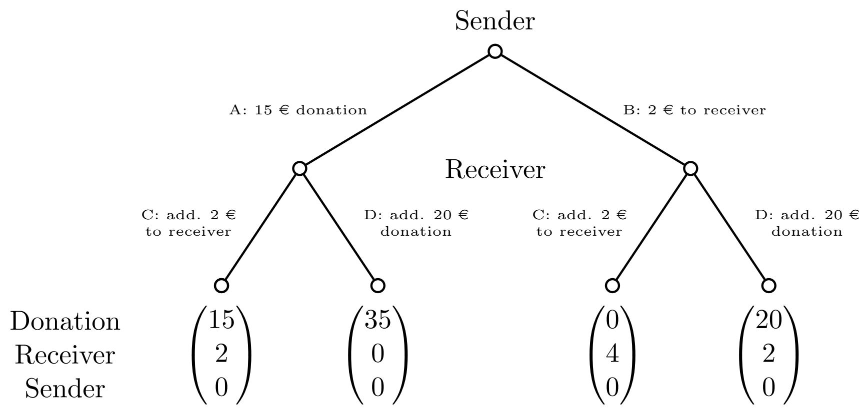
***Aligned* version.** The sender can either send a message that Option A triggers the higher donation (truth) or a message that Option B will trigger it (lie). Because almost all subjects follow this message, just as in the main game, sending the truthful message achieves the more desirable consequence. Accordingly, both moral principles prescribe sending this message.

B.3 Bribe game

Two subjects are paired together, one playing the role of a sender and the other playing a receiver. Payoffs for both players and the charity are determined by the sender's choices, knowing the receiver's predetermined conditional responses.

The choice task unfolds in two stages, summarized by Figure B.2. In the first stage, the sender chooses between Option A, which implements a 15 Euros donation to the charity, and Option B, which pays 2 Euros to the receiver. In the second stage, knowing what the sender chose, the receiver decides between Option C, which yields them 2 Euros, and Option D, which triggers a separate 20 Euros donation to the charity.

Figure B.2: Bribe game



Receivers' contingent decisions are elicited using the strategy method, and all senders are (truthfully) informed that their paired receiver was one of those who decided to make their choice between

[30]We obtained this number through a pilot.

C and D hinge on their sender's decision, as follows:

- If the sender chooses A, the receiver will choose C and take 2 Euros for themselves.
- If the sender chooses B, the receiver will choose D, thus triggering a donation of 20 Euros to charity while keeping the 2 Euros sent by the other player.

Thus, the receiver, in effect, demands a bribe in order to implement the most socially desirable option (the word bribe is never used in the instructions). Again, we are only interested in the choices of the senders confronted with such a demand, who constitute the subjects of the main experiment. Receivers are part of a separate sample, and their choices are not of primary interest.[31]

Tradeoff. Note that the sender is not facing a strategic situation but, once again, a simple dilemma pitting a deontological approach, which entails sticking with Option A and making a donation of 15 Euros without regard to ultimate consequences, against a consequentialist one that will increase the total donation to 20 Euros but requires paying the bribe demanded by the other player in order to achieve this end.

***Aligned* version.** Instead of being matched with receivers who demand a bribe (choice of B over A) for donating (choosing D over C), senders are matched with receivers who demand that they donate (Option A) in order for themselves to choose their donation option (Option D). To implement the socially most desirable option, senders must thus choose to donate rather than bribe; consequentialist and deontological prescriptions coincide.

B.4 Group donation game

For this task, adapted from Falk, Neuber, and Szech (2020), subjects are sorted into groups of size 6. Each member makes their decision autonomously from the others, but the consequences of the actions taken by any can affect everyone. At the start of the game, the group is entrusted with 15 Euros, which is preset to be donated to the children's cancer charity by the end of the study. The first five members (first-movers), who take action earliest, are part of a separate sample, while the sixth one (second-mover) is part of the main study. First-movers simultaneously choose between:

- Option A: This choice has no further consequence, but simply preserves the donation intact if it is still relevant.
- Option B: This choice grants the member who chooses it 2 Euros as an additional payment for themselves. If even just one group member chooses this option, however, the group's donation will be canceled.

The second-mover also has two choices: Option A entails no additional payment to anyone and preserves the donation in case it is still intact; Option B grants 2 Euros to some subject outside the group, but also destroys the donation in case it is still intact.

[31] Among them, 60% demanded a bribe to choose the donation, while 40% chose either the donation (23%) or the money (17%) unconditionally. All senders were paired with one among the first group.

Before making their decision, second-movers learn whether the donation has already been destroyed by the choices of the first movers. Importantly, in our experiment, at least one first mover in every group opted for option B. As in Falk, Neuber, and Szech (2020), all second movers thus choose with full knowledge that the donation had, effectively, already been canceled.

Tradeoff. Option A is thus entirely inconsequential, but choosing it allows the subject to (anonymously) express a moral preference for donating, an act aligned with deontological principles (such as the categorical imperative: one should will that everyone always chooses Option A). Choosing Option B, meanwhile, is clearly consequentialist, as it generates a positive outcome for another participant who took no part in the task, without affecting the already foregone charitable donation.

***Aligned* version.** A group of 6 subjects is now entrusted with a 2 Euro payment destined for another subject who is not a group member. As in the main experiment, Option A has no further consequence but simply preserves that payment, if still intact. If anyone chooses Option B, on the other hand, the 2 Euro payment is canceled and a 15 Euro donation to the charity is triggered (each time Option B is chosen). As in the main game, our subjects of interest are second-movers, who decide last and are informed about whether the 2 Euro payment has been destroyed by previous choices. Again, in all groups, at least one first mover chose Option B, thus rendering Option A for the second mover inconsequential. As a consequence, both moral principles are aligned in their prescription of Option B, as this option is both consequential and morally desirable under deontological principles.

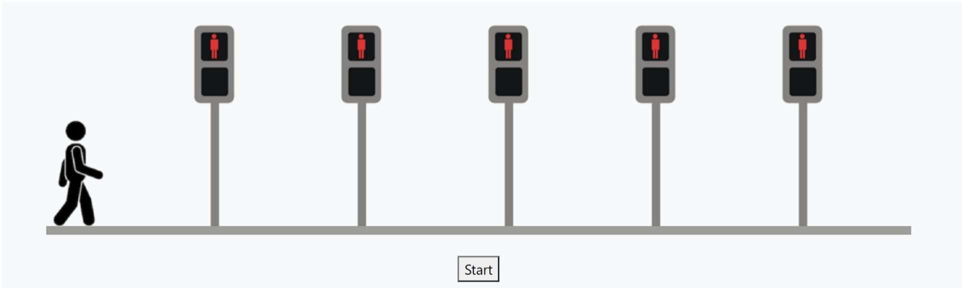
B.5 Statement choice

Subjects first receive information on the harmful effects of CO_2 on the environment and its contribution to climate change. They are subsequently informed that, as part of the study, the purchase of carbon offsets has been prearranged, each such certificate corresponding to offsetting 1 ton of CO_2 from the atmosphere. It is also explained to them that once they have completed their task, the computer will randomly destroy (i.e., not follow through with the purchase) one certificate with a probability of 50%. The destruction of the certificate thus means that the planned removal of 1 ton of CO_2 will not take place. This process is completely independent of any subject's actions, and the certificate feature is included only to make more salient both the climate-change problem and the fact that there exist ways to alleviate it. In this context, subjects can choose to submit one of the following declarations (pressing the corresponding button):

1. "I support the preservation and protection of the environment."
2. "I support the destruction of the environment."

As explained to the subjects, the first statement leads to another subject receiving 2 Euros, whereas the latter triggers a donation of 15 Euros to the children's cancer charity. In addition to being fully anonymous (like all choices in our experiments), subjects are informed that the results of the experiment will not be used for any other purpose, such as an opinion poll, thus depriving the statements of any instrumental value.

Figure B.3: Decision screen of the rule-following task



Tradeoff. The dilemma is thus to stick to one's values and submit the first statement,[32] as at least strongly suggested by the deontological approach, or to submit the second one in order to achieve an unambiguously better outcome, in line with a consequentialist view. Examining people's willingness to (anonymously) make a statement that is antithetical to their moral identity is a procedure similar to that in Bursztyn et al. (2020), but in our case, the consequence of doing so is not a material reward for oneself (one fifth or a day's wage in that paper), but once again creating a positive social externality.

***Aligned* version.** Here, choosing the statement in support of the environment triggers the donation, while choosing the one supporting the destruction of the environment leads to another subject receiving 2 Euros. Hence, both moral principles prescribe endorsing the first statement.

B.6 Rule-following task

The design is taken from Kimbrough and Vostroknutov (2016). The subject controls a stick figure walking across the computer screen along a series of traffic lights and decides how long to wait at each, see Figure B.3 for a visualization. Initially, the figure is at the left of the screen, and all lights are red. Once the animation starts, the figure "walks" towards the end of the path (right of the screen), automatically stopping and waiting at every red light. Each time, however, the subject can decide to press a button that causes the figure to proceed through the red light without waiting.

Subjects receive an endowment of 8 Euros and incur a deduction of 0.08 Euros for each second it takes the figure to walk across the screen. Without stopping at any red light, it takes four seconds to complete the track, costing the subject about 2 Euros in total. Waiting at each red light roughly doubles these losses to about 4 Euros. In the instructions received by subjects, they are told that the "rule of the game" is to stop at each red light until it turns green; there is no enforcement or incentive to follow the rule, however.

Tradeoff. The design creates a tradeoff between a deontological approach to the problem, which entails following the stated rule of stopping at each traffic light (or the meta-rule that "one should play by the rules") and incurring losses, versus a consequentialist calculus, which favors maximizing the total payoff by breaking the non-instrumental rule (walking through red lights).

[32]As stated in the main text, our sample consists of very environmentally-conscious subjects: 96% agree that fighting climate change is important, and 99% agree that measures to protect the environment are important.

***Aligned* version.** Subjects now start with 0 Euro and receive 0.08 for each second that the figure is walking across the screen, up to a maximum of 8 Euros. Accordingly, following the rule by waiting at the traffic lights maximizes the payoff that subjects can receive from the game, and is thus the aligned choice.

C Details on self-versus-others choice tasks

This section provides further details on the self versus other choices covered in Section 3.2.

C.1 Dictator game

Subjects play two variations of a dictator game, a giving and a taking game. In each, they can allocate 20 Euros between themselves and a charity helping children suffering from cancer. In the taking game, 20 Euros 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 Euros and can donate an amount to the charity.

C.2 Trust game

Subjects play both roles of the standard trust game with an endowment of 5 Euros and a multiplication factor of 3. Senders can send integers from 0 to 5 Euros. 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.

C.3 Public goods game

Subjects play the standard public goods game with an endowment of 5 Euros, a group size of 3, and a marginal per capita return of 0.5.

C.4 Moral-luck game

There are two players: a first-mover who chooses between two lotteries with different payoff distributions for themselves and for the charity, and a dictator who allocates additional money between themselves and the first-mover. These features make the game one of conditional altruism, which is why it is included in this block. At the same time, it is designed to measure and compare the extent to which dictators reward socially desirable intentions versus socially desirable outcomes, meaning the game could also be informative about deontological versus consequentialist preferences, or norms.

First-movers, whose behavior is not *per se* the object of interest, choose between two lotteries: (i) Lottery M, which yields 10 Euros to self with 70% probability and a 15 Euros donation to the charity with 30% probability; (ii) Lottery D, which yields 10 Euros to self with 30% probability and a 15 Euros donation to the charity with 70% probability.

Subjects in the role of the dictator then make, using the strategy method, allocation choices to the first-mover they are paired with, for each of the four possible choice-outcome:

S1 First-mover chooses M, 10 Euros payment realizes

S2 First-mover chooses D, 10 Euros payment realizes

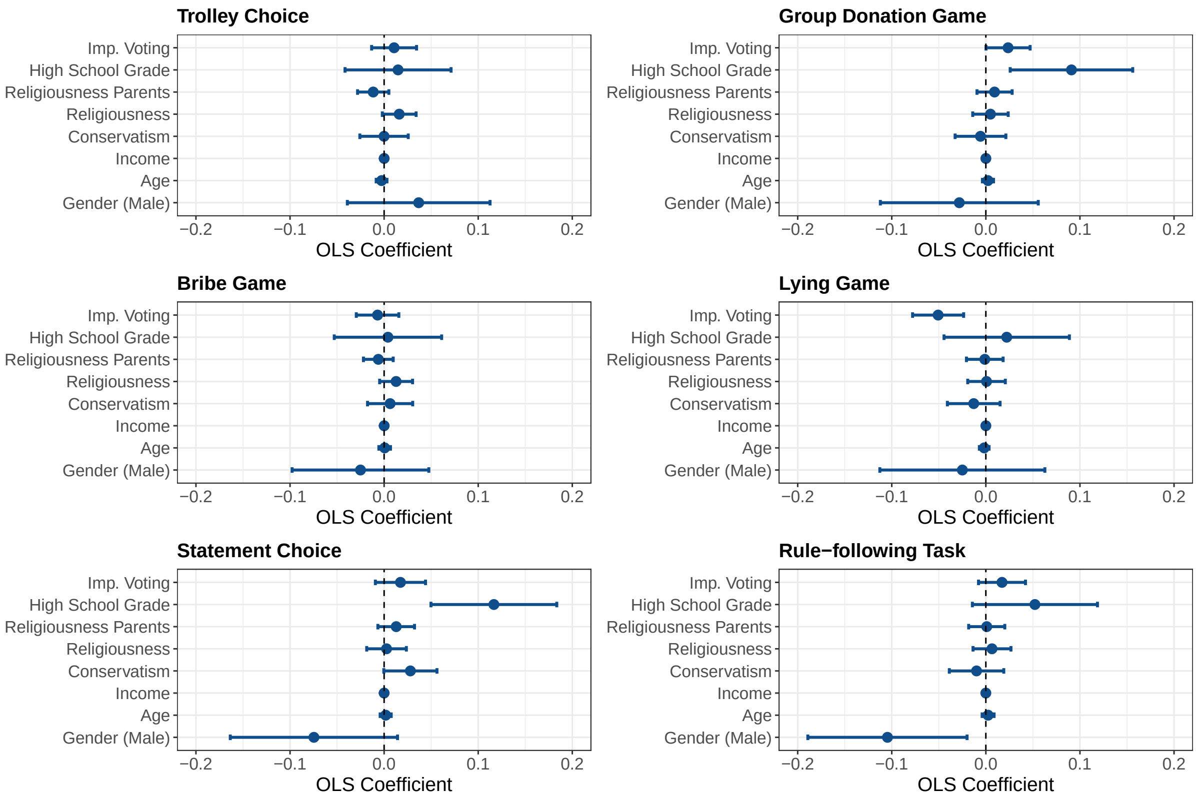
S3 First-mover chooses M, 15 Euros donation realizes

S4 First-mover chooses D, 15 Euros donation realizes

The dictator is endowed with 10 Euros and the amount x they allocate to the first mover is tripled (they keep the remaining $10 - x$), in order to induce positive giving in each of the four possible situations. The first-mover thus potentially receives money from the allocation task and the lotteries. The dictator's choices then allow us to ask: (i) fixing the lottery outcome, to what extent do dictators take the first-mover's lottery choice ("intention") into account when allocating money to them? This is our variable *rewarding intentions*, defined as the amount allocated to the first mover in $S4 - S3$. (ii) fixing the lottery choice, to what extent do dictators take the outcome into account in their allocation? This is our variable *rewarding consequences*, defined as the amount allocated to the first mover in $S4 - S2$. (iii) is their decision more responsive to the first-mover's *ex-ante* lottery choice or to its *ex-post* realization? This is our variable *rewarding consequences - intention*, defined as the amount allocated to the first mover in $(S4 + S3) - (S2 + S1) - ((S4 - S3) + (S2 - S1))$.

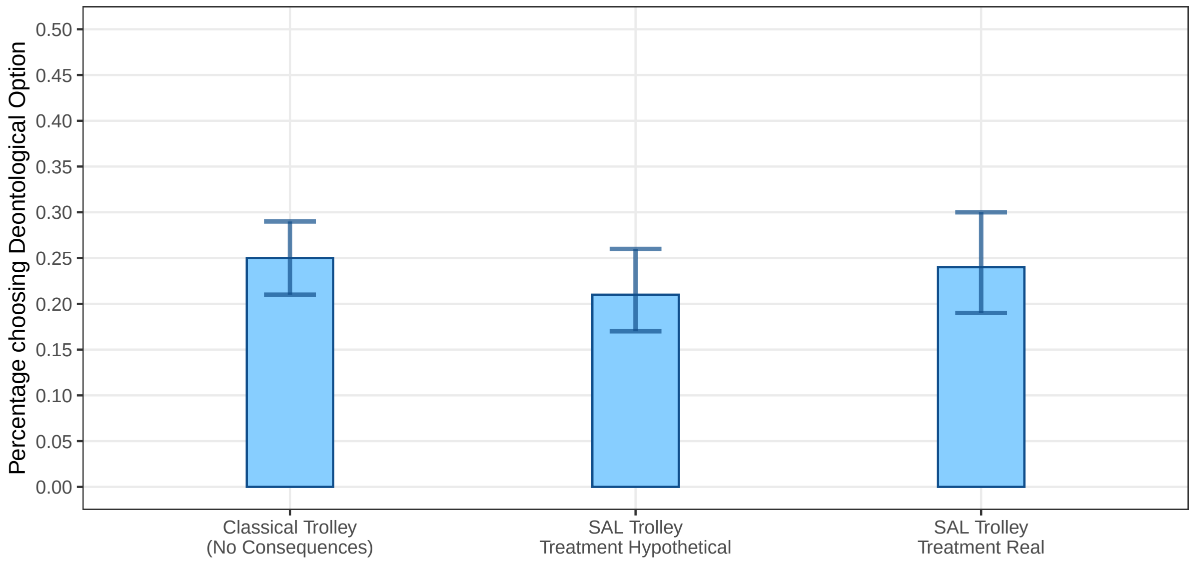
D Additional figures

Figure D.1: The association of demographic variables with EVM choice behavior



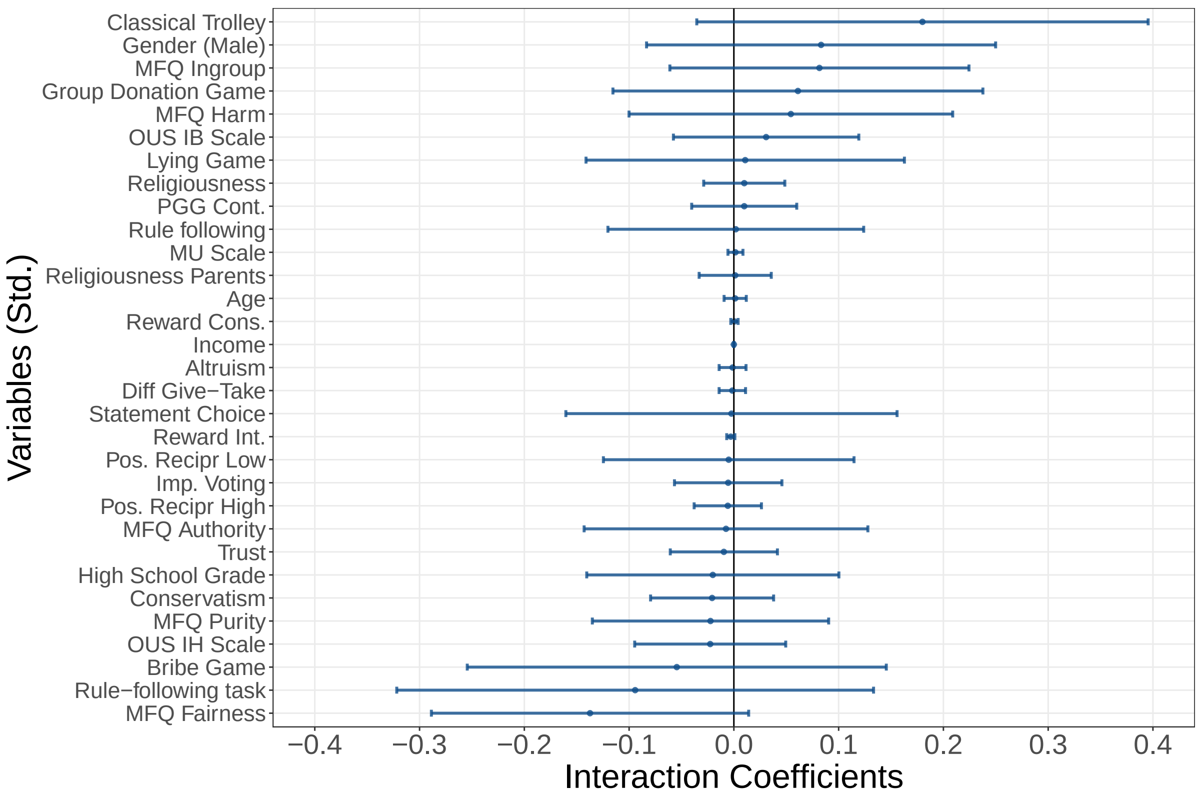
Notes: The graph displays coefficients obtained from an OLS regression where the dependent variable is EVM behavior in the corresponding choice task as dependent variable, with the variable equal to one if the deontological option was chosen and zero if the consequentialist option was chosen. Bars indicate 95% confidence intervals constructed from robust standard errors.

Figure D.2: The effect of changing realism and consequences on behavior in the trolley dilemma among high comprehension subjects



Notes: Each bar displays the fraction of subjects choosing the deontological option - not switching the lever - in the trolley dilemma, using only subjects that make no more than one mistake in the trolley comprehension quiz. The first two bars display behavior in the classical train track trolley, the last two behavior using the Saving A Life paradigm. Treatment *Real* indicates the treatment in which the SAL trolley has real consequences, in treatment *Hypothetical* decisions are without consequences. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

Figure D.3: Differences in characteristics of subjects between the *Hypothetical* and *Real* treatments



Notes: The graph displays interactions term coefficients obtained from an OLS regression where the dependent variable is an indicator that equals one for observations from the *SAL-Real* treatment and zero for observations from the *SAL-Hypothetical* treatment. Bars indicate 95% confidence intervals constructed from robust standard errors.

E Robustness experiment

This section describes a robustness experiment addressing the concern that the probabilistic implementation employed in the treatment *Real* introduced an insufficient real consequence incentive. Specifically, in treatment *Real*, each choice in the Saving a Life trolley has a 10% chance of being implemented with real consequences. In comparison, in treatment *Hypothetical*, all choices had no real consequences. We find that subjects' behavior does not differ between the two treatments. One potential explanation is that subjects perceive a 10% implementation probability as essentially hypothetical, rendering the treatment comparison problematic. In the robustness experiment, we show that subjects are highly responsive to a chance from 0% implementation probability to a positive probability in self-versus-other tradeoffs. We use the design and results of one treatment of Bénabou et al. (2023), to which we add a hypothetical treatment.

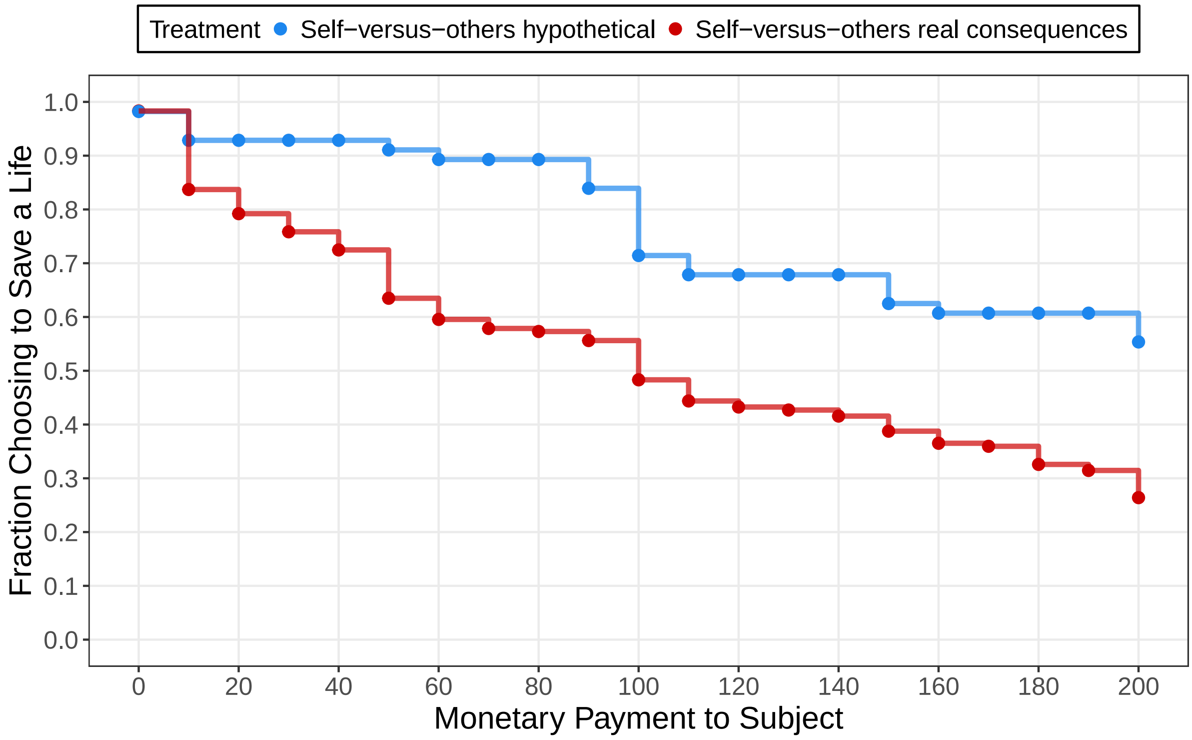
Design. We used the same Saving a Life paradigm as employed in the main experiment. However, instead of the trolley's ends-versus-means tradeoff, we employed a self-versus-other tradeoff: subjects could choose whether to trigger a donation sufficient to save one human life or take money for themselves as additional payment. Subjects faced this tradeoff multiple times in the form of a price-list design. In 21 contingent choices starting with 0 Euro and proceeding in 10 Euros increments up to 200 Euros, subjects could indicate whether they want to save a life or take the respective money for themselves.

Treatment. There were two between-subject treatments. In the *self-versus-other real* treatment[33], subjects' choices had a chance to be implemented with real consequences. Out of each session (consisting of 20-24 subjects), two subjects were drawn. For these selected subjects, one price from the price list was randomly drawn, and their pre-stated choices for the drawn price were implemented. Therefore, in this treatment, with a 10% chance, subjects either triggered the donation or received up to 200 Euros through their decisions. We subsequently ran another treatment, called *self-versus-other hypothetical*, in which subjects' choices were purely hypothetical. In total, 178 subjects took part in the *self-versus-other real* treatment (Bénabou et al., 2023), and we recruited 56 subjects for the treatment *self-versus-other hypothetical*.

Results. Introducing real consequences to this self-versus-other tradeoff resulted in significantly different choice distributions ($p < 0.001$, Kolmogorov-Smirnov test). As displayed in Figure E.1, for all prices except 0 and 10 Euros, real consequences lead to a significant decrease in the likelihood that subjects choose to save a life instead of taking the offered amount. For instance, at 100 Euros, 71% of subjects choose to save a life when consequences are hypothetical, while only 48% do so when choices might have real consequences ($p = 0.003$, two-sided Fisher's exact test). Consequently, introducing real consequences through a probabilistic payment system leads to markedly different results in self-versus-other tradeoffs.

[33] In Bénabou et al. (2023), this treatment is called *MPL Low Image*.

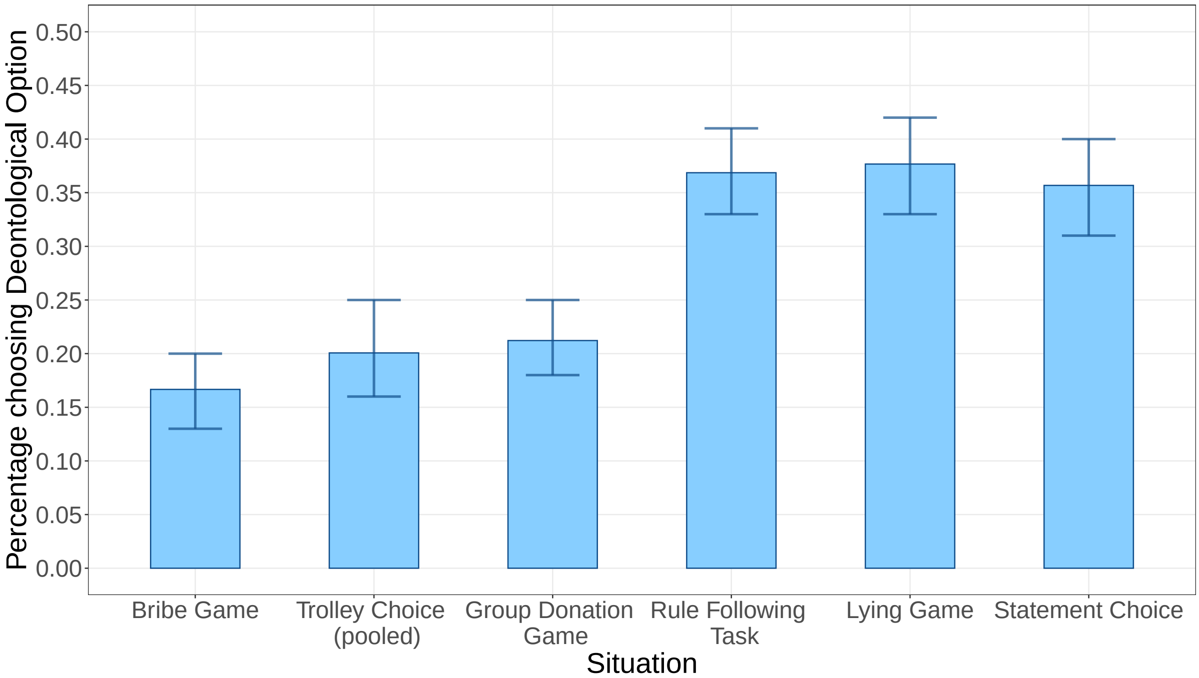
Figure E.1: The effect of changing consequences on behavior in self versus other tradeoff situations



Notes: The figure displays the fraction of subjects choosing to save a life for each offered price separately for the *self-versus-other hypothetical* and *self-versus-other real consequences* treatments.

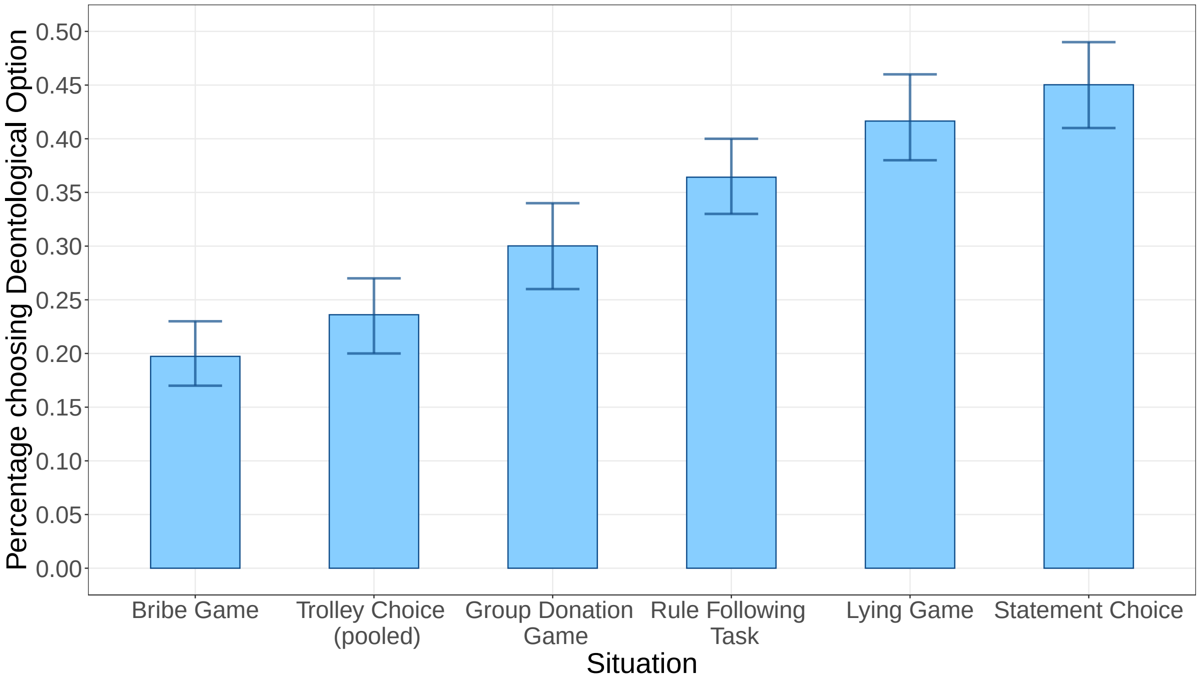
F Robustness analysis ends-versus-means decisions

Figure F.1: Distribution of decisions in ends-versus-means choice tasks among high comprehension subjects



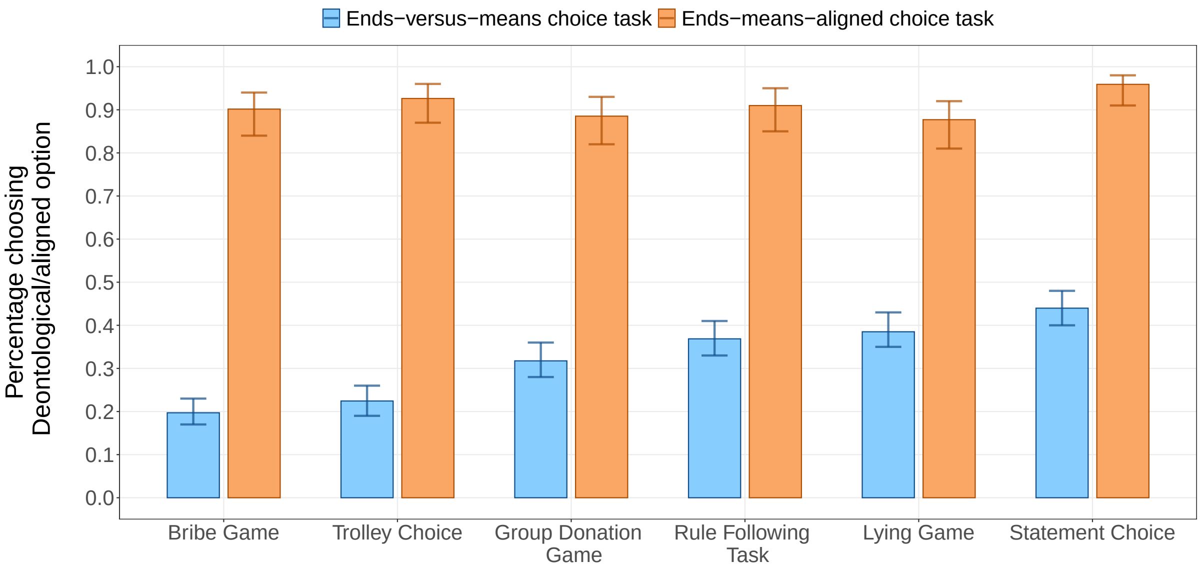
Notes: Each bar displays the fraction of subjects choosing the deontological option in ends-versus-means choice tasks, using only subjects who answer all comprehension and recall questions correctly. See Section 3.1 for details on the situations and the definition of the deontological option. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

Figure F.2: Distribution of decisions in ends-versus-means choice tasks using the full sample of subjects



Notes: Each bar displays the fraction of subjects choosing the deontological option in ends-versus-means choice tasks, using the full sample of 593 subjects. See Section 3.1 for details on the situations and the definition of the deontological option. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

Figure F.3: Distribution of decisions in the ends-versus-means and ends-means-aligned choice tasks



Notes: Each bar displays the fraction of subjects choosing the deontological option in ends-versus-means choice tasks (left bar in blue) and the fraction choosing the aligned choice in ends-means-aligned choice tasks (right bar in orange). See Section 3.1 for details on the situations and the definition of the deontological and aligned option. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

G Further robustness analyses

Table G.1: Correlations of ends-versus-means, self-versus-other, and questionnaire measures

	Ends-versus-means choice tasks						Self-versus-others choice tasks									Questionnaires								
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity
SAL Trolley		0.05	0.05	0.06	0.16	0.03	-0.01	-0.03	0.02	-0.05	-0.03	0.00	-0.02	0.05	-0.02	0.50	-0.12	-0.29	-0.04	-0.06	0.00	-0.06	0.00	0.05
Group Don.			0.14	0.15	-0.09	-0.06	0.05	0.03	-0.04	0.14	0.05	-0.02	0.09	-0.03	-0.02	0.02	0.09	0.02	-0.02	0.10	0.02	0.08	0.03	0.12
Bribe				0.19	0.04	-0.08	-0.04	-0.02	-0.11	-0.06	-0.15	-0.06	-0.06	-0.18	-0.18	-0.01	-0.03	-0.04	0.01	0.02	-0.04	0.08	0.05	0.14
Statement Choice					-0.06	0.04	0.00	0.01	-0.12	0.04	-0.07	0.00	0.05	-0.15	-0.04	-0.01	0.03	-0.07	-0.02	0.03	-0.03	0.15	0.13	0.17
Lying						0.00	-0.03	-0.04	-0.03	-0.03	0.00	-0.05	0.00	-0.02	-0.01	0.14	-0.15	-0.11	-0.05	-0.03	-0.02	-0.08	-0.04	-0.10
Rule following							0.11	0.06	0.03	0.11	0.10	0.03	0.11	-0.01	0.13	-0.02	0.08	0.00	-0.07	0.04	0.00	-0.08	-0.02	-0.02
Altruism								0.65	0.20	0.31	0.38	0.40	0.36	0.23	0.23	0.00	0.22	0.00	-0.13	0.19	0.17	0.01	-0.02	0.00
Altruism Taking									0.16	0.26	0.35	0.36	0.28	0.21	0.21	0.01	0.20	0.00	-0.13	0.20	0.15	0.00	-0.02	-0.04
Trust										0.35	0.45	0.47	0.38	0.30	0.23	0.00	0.05	0.02	-0.13	0.00	0.06	-0.06	-0.13	-0.17
Pos. Recipr. Low											0.71	0.36	0.35	0.13	0.14	-0.02	0.17	0.03	-0.16	0.08	0.09	-0.02	-0.08	-0.04
Pos. Recipr. High												0.45	0.41	0.30	0.21	0.01	0.18	0.04	-0.18	0.09	0.11	-0.12	-0.16	-0.15
PGG Cont.													0.40	0.31	0.22	0.03	0.11	-0.13	-0.26	0.06	0.12	-0.02	-0.07	-0.09
Reward Cons.- Int.														0.24	0.55	-0.02	0.08	-0.01	-0.15	0.06	0.07	-0.07	-0.12	-0.11
Reward Int.															0.40	0.07	0.02	0.00	-0.12	-0.01	0.04	-0.10	-0.08	-0.14
Reward Cons.																0.04	0.01	0.01	-0.04	0.05	0.06	-0.09	-0.08	-0.09
Classical Trolley																	-0.21	-0.43	-0.04	-0.04	0.01	-0.08	-0.06	-0.01
OUS IB																		0.23	-0.06	0.35	0.24	0.08	-0.04	0.05
OUS IH																			0.15	-0.03	-0.03	0.16	0.15	0.07
MU Scale																				0.02	0.02	0.22	0.16	0.20
MFQ Harm																					0.58	0.23	0.17	0.26
MFQ Fairness																						0.16	0.08	0.14
MFQ Ingroup																							0.57	0.51
MFQ Authority																								0.55
MFQ Purity																								

Notes: Colors in heatmap change with each 0.1 correlation magnitude.

Table G.2: Correlations of ends-versus-means, self-versus-other and questionnaire measures comprehension sample

	Ends-versus-means choice tasks						Self-versus-others choice tasks									Questionnaires								
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.-Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity
SAL Trolley		0.00	-0.02	0.05	0.17	0.10	-0.01	0.00	0.01	-0.07	-0.05	0.00	0.00	0.03	-0.03	0.56	-0.13	-0.30	0.00	-0.07	-0.05	-0.04	0.02	0.06
Group Don.			0.08	0.00	-0.14	-0.05	0.03	0.05	0.06	0.13	0.08	0.01	0.14	0.04	0.07	0.02	0.11	0.03	-0.01	0.10	0.00	0.02	-0.04	0.06
Bribe				0.09	0.04	-0.08	0.00	0.01	-0.07	-0.01	-0.08	-0.07	-0.09	-0.08	-0.11	0.01	-0.07	0.07	0.00	-0.06	-0.07	0.03	0.00	0.06
Statement Choice					-0.08	0.07	0.03	0.05	-0.13	0.02	-0.03	0.00	0.03	-0.12	-0.06	0.01	0.04	-0.15	-0.08	0.03	0.05	0.09	0.06	0.10
Lying						0.04	-0.05	-0.09	-0.04	-0.04	0.01	-0.03	0.03	-0.03	0.03	0.15	-0.22	-0.10	-0.05	-0.06	-0.05	-0.05	-0.02	-0.04
Rule following							0.08	0.04	-0.01	0.09	0.10	0.04	0.10	0.01	0.11	-0.02	0.05	-0.05	-0.12	0.00	-0.03	-0.14	-0.06	-0.07
Altruism								0.66	0.16	0.32	0.32	0.43	0.30	0.26	0.20	-0.01	0.24	0.01	-0.16	0.21	0.16	-0.05	-0.02	0.00
Altruism Taking									0.11	0.25	0.34	0.37	0.25	0.27	0.22	0.00	0.26	0.01	-0.15	0.23	0.12	-0.06	-0.01	-0.04
Trust										0.32	0.42	0.46	0.31	0.31	0.18	-0.01	0.06	0.06	-0.10	-0.04	0.05	-0.01	-0.10	-0.16
Pos. Recipr. Low											0.72	0.40	0.33	0.21	0.18	-0.04	0.16	0.00	-0.18	-0.04	0.05	-0.06	-0.12	-0.07
Pos. Recipr. High												0.46	0.37	0.33	0.19	0.00	0.18	0.06	-0.20	0.03	0.02	-0.13	-0.16	-0.18
PGG Cont.													0.39	0.38	0.27	0.03	0.11	-0.10	-0.23	0.03	0.08	-0.06	-0.06	-0.11
Reward Cons.- Int.														0.35	0.61	0.03	0.12	-0.03	-0.18	0.02	0.03	-0.13	-0.12	-0.10
Reward Int.															0.42	0.06	0.07	0.01	-0.17	0.02	0.01	-0.10	-0.07	-0.12
Reward Cons.																0.09	0.04	-0.05	-0.04	0.05	0.07	-0.09	-0.13	-0.09
Classical Trolley																	-0.22	-0.46	0.00	-0.06	-0.04	-0.06	-0.05	0.01
OUS IB																		0.24	-0.05	0.38	0.22	0.07	-0.02	0.01
OUS IH																			0.19	-0.01	-0.03	0.17	0.14	0.04
MU Scale																				0.09	0.05	0.24	0.14	0.20
MFQ Harm																					0.51	0.22	0.17	0.26
MFQ Fairness																						0.18	0.08	0.11
MFQ Ingroup																							0.55	0.50
MFQ Authority																								0.53
MFQ Purity																								

Notes: Colors in heatmap change with each 0.1 correlation magnitude.

Table G.3: Correlations of ends-versus-means, self-versus-other and questionnaire measures memory check sample

	Ends-versus-means choice tasks						Self-versus-others choice tasks									Questionnaires								
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.-Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity
SAL Trolley		0.05	0.01	0.05	0.16	0.04	0.05	0.00	0.05	0.00	0.02	0.05	0.02	0.10	0.00	0.54	-0.09	-0.31	-0.07	-0.08	0.02	-0.02	-0.01	0.06
Group Don.			0.07	0.07	-0.07	0.00	0.12	0.12	0.04	0.19	0.11	-0.01	0.14	0.04	0.03	0.03	0.06	0.03	-0.03	0.08	0.04	-0.02	-0.05	0.04
Bribe				0.13	0.08	-0.09	-0.01	0.05	-0.05	-0.02	-0.14	-0.06	-0.05	-0.15	-0.12	-0.02	-0.02	-0.10	0.02	0.12	0.05	0.10	0.02	0.05
Statement Choice					-0.05	0.04	0.04	0.08	-0.04	0.06	-0.02	0.09	0.09	-0.07	0.04	0.01	0.04	-0.07	-0.11	0.09	0.09	0.13	0.15	0.13
Lying						0.07	-0.06	-0.10	-0.04	0.03	0.00	-0.05	0.01	-0.03	-0.01	0.11	-0.14	-0.11	-0.10	-0.07	-0.02	-0.08	-0.05	-0.10
Rule following							0.13	0.11	0.01	0.10	0.15	0.04	0.16	0.01	0.11	0.00	0.11	0.00	-0.08	0.11	0.08	-0.12	-0.02	0.01
Altruism								0.65	0.18	0.32	0.38	0.40	0.35	0.30	0.22	0.02	0.28	-0.05	-0.10	0.24	0.25	-0.06	-0.02	-0.01
Altruism Taking									0.12	0.28	0.38	0.33	0.29	0.24	0.24	0.01	0.27	-0.04	-0.08	0.23	0.18	-0.02	0.00	-0.01
Trust										0.31	0.41	0.45	0.39	0.24	0.21	0.01	0.05	0.01	-0.08	0.03	0.08	0.01	-0.13	-0.09
Pos. Recipr. Low											0.72	0.34	0.38	0.12	0.16	-0.01	0.16	0.02	-0.13	0.13	0.12	-0.02	-0.05	-0.02
Pos. Recipr. High												0.46	0.48	0.29	0.25	0.03	0.21	0.03	-0.14	0.15	0.16	-0.13	-0.14	-0.10
PGG Cont.													0.44	0.34	0.25	0.02	0.14	-0.15	-0.21	0.06	0.11	0.00	-0.06	-0.07
Reward Cons.- Int.														0.33	0.56	-0.02	0.16	-0.01	-0.12	0.12	0.13	-0.11	-0.12	-0.08
Reward Int.															0.46	0.13	0.09	0.00	-0.07	0.03	0.05	-0.05	-0.08	-0.08
Reward Cons.																0.07	0.08	0.03	-0.01	0.08	0.09	-0.14	-0.11	-0.07
Classical Trolley																	-0.25	-0.47	-0.05	-0.08	0.02	-0.09	-0.05	0.02
OUS IB																		0.23	-0.08	0.40	0.24	0.06	-0.06	0.02
OUS IH																			0.17	-0.03	-0.04	0.14	0.13	0.05
MU Scale																				-0.02	-0.01	0.19	0.14	0.20
MFQ Harm																					0.54	0.16	0.12	0.26
MFQ Fairness																						0.08	0.00	0.12
MFQ Ingroup																							0.54	0.48
MFQ Authority																								0.52
MFQ Purity																								

Notes: Colors in heatmap change with each 0.1 correlation magnitude.

Table G.4: Correlations of ends-versus-means, self-versus-other and questionnaire measures excluding speeders

	Ends-versus-means choice tasks						Self-versus-others choice tasks									Questionnaires								
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity
SAL Trolley		0.03	0.06	0.06	0.17	0.06	-0.01	-0.03	0.02	-0.05	-0.03	-0.02	-0.02	0.07	-0.01	0.52	-0.15	-0.29	-0.02	-0.07	-0.03	-0.06	-0.01	0.06
Group Don.			0.13	0.14	-0.06	-0.08	0.07	0.04	-0.04	0.13	0.05	-0.01	0.07	-0.02	-0.01	0.01	0.09	0.03	0.00	0.10	0.03	0.09	0.02	0.09
Bribe				0.16	0.06	-0.09	0.01	0.03	-0.09	-0.05	-0.12	-0.07	-0.02	-0.16	-0.14	-0.01	-0.03	-0.04	0.03	0.06	0.00	0.09	0.06	0.16
Statement Choice					-0.04	0.05	0.04	0.05	-0.12	0.04	-0.08	0.00	0.07	-0.16	0.00	0.02	0.03	-0.08	-0.01	0.08	0.01	0.17	0.15	0.18
Lying						0.02	-0.03	-0.05	-0.05	-0.02	-0.01	-0.05	0.01	-0.02	-0.02	0.15	-0.17	-0.10	-0.06	-0.06	-0.03	-0.08	-0.03	-0.09
Rule following							0.10	0.06	0.00	0.10	0.09	0.00	0.08	-0.03	0.11	-0.01	0.07	0.00	-0.07	0.02	-0.03	-0.08	0.01	-0.03
Altruism								0.64	0.16	0.30	0.35	0.40	0.33	0.22	0.18	-0.03	0.21	0.01	-0.13	0.19	0.17	0.04	0.02	0.02
Altruism Taking									0.12	0.25	0.33	0.37	0.26	0.18	0.17	-0.03	0.19	0.01	-0.12	0.20	0.14	0.02	0.01	-0.04
Trust										0.32	0.42	0.46	0.36	0.27	0.21	-0.01	0.01	0.05	-0.12	-0.03	0.06	-0.04	-0.11	-0.17
Pos. Recipr. Low											0.69	0.33	0.32	0.11	0.11	-0.03	0.16	0.04	-0.14	0.06	0.07	-0.03	-0.08	-0.05
Pos. Recipr. High												0.43	0.39	0.28	0.16	0.00	0.17	0.07	-0.16	0.06	0.10	-0.14	-0.15	-0.16
PGG Cont.													0.40	0.31	0.23	0.01	0.12	-0.10	-0.26	0.09	0.13	-0.01	-0.05	-0.06
Reward Cons.- Int.														0.22	0.54	-0.02	0.06	-0.01	-0.13	0.03	0.05	-0.09	-0.11	-0.11
Reward Int.															0.39	0.08	0.00	-0.01	-0.10	-0.04	0.04	-0.11	-0.07	-0.16
Reward Cons.																0.05	-0.01	0.01	-0.01	0.02	0.04	-0.10	-0.06	-0.10
Classical Trolley																	-0.23	-0.44	-0.02	-0.03	0.02	-0.10	-0.06	-0.02
OUS IB																		0.25	-0.06	0.35	0.25	0.08	-0.04	0.02
OUS IH																			0.16	-0.03	-0.04	0.16	0.15	0.06
MU Scale																				0.02	0.03	0.23	0.15	0.20
MFQ Harm																					0.60	0.24	0.16	0.24
MFQ Fairness																						0.18	0.08	0.15
MFQ Ingroup																							0.56	0.51
MFQ Authority																								0.54
MFQ Purity																								

Notes: Colors in heatmap change with each 0.1 correlation magnitude.

G.1 Using dichotomous self-versus-other variables

The variables in the EVM block are dichotomous, while those in the SVO block are more fine-grained. To analyze how this differences affects our comparison of correlations, we dichotomize the SVO variables in two different ways. First, we split subjects based on whether they fully behave according to the classical model of maximizing self-interest. That is, if a subject donates nothing to the charity, does not cooperate at all in the public goods game, sends nothing in the trust game, etc., we code the respective variable as zero, and one otherwise. Second, we categorize subjects based on a median split for each respective behavior. For example, the altruism variable is set to zero if a subject donates less than the median level, and one if they donate more. For both approaches, we again find a high degree of consistency within the SVO block, and no correlation with the ends-versus-means block; see Table G.5 for the first approach, and Table G.6 for the second.

Table G.5: Correlations between ends-versus-means and self-versus-other blocks using binary categorization for the later variables

	Ends-versus-means choice tasks						Self-versus-others choice tasks								
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons. comp Int.	Reward Int.	Reward Cons.
SAL Trolley		0.05	0.05	0.06	0.16	0.03	-0.04	-0.09	-0.03	-0.09	-0.04	-0.04	-0.07	-0.06	-0.06
Group Don.			0.14	0.15	-0.09	-0.06	0.05	0.06	-0.03	0.06	0.07	0.01	0.10	0.10	0.10
Bribe				0.19	0.04	-0.08	-0.02	-0.03	-0.08	-0.09	-0.08	-0.04	-0.12	-0.12	-0.12
Statement Choice					-0.06	0.04	0.05	0.07	-0.04	0.03	0.02	0.03	0.01	0.02	0.00
Lying						0.00	-0.06	-0.08	-0.04	-0.01	0.00	-0.05	-0.05	-0.05	-0.05
Rule following							0.07	0.06	0.06	0.07	0.09	0.02	0.04	0.05	0.05
Altruism								0.66	0.25	0.35	0.40	0.40	0.39	0.40	0.40
Altruism Taking									0.27	0.34	0.35	0.42	0.39	0.40	0.40
Trust										0.45	0.44	0.43	0.42	0.41	0.42
Pos. Recipr. Low											0.78	0.44	0.52	0.51	0.51
Pos. Recipr. High												0.45	0.55	0.54	0.54
PGG Cont.													0.49	0.48	0.48
Reward Cons.- Int.														0.98	0.99
Reward Int.															0.97
Reward Cons.															

Table G.6: Correlations between ends-versus-means and self-versus-other blocks using median splits for the later variables

	Ends-versus-means choice tasks						Self-versus-others choice tasks								
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons. comp Int.	Reward Int.	Reward Cons.
SAL Trolley		0.05	0.05	0.06	0.16	0.03	-0.02	-0.01	-0.02	-0.08	-0.02	0.01	-0.03	-0.03	-0.07
Group Don.			0.14	0.15	-0.09	-0.06	0.02	0.05	-0.03	0.13	0.00	-0.02	0.09	-0.04	-0.03
Bribe				0.19	0.04	-0.08	0.00	0.01	-0.10	-0.06	-0.14	-0.03	-0.07	-0.16	-0.20
Statement Choice					-0.06	0.04	0.01	0.01	-0.10	0.05	-0.08	0.01	0.05	-0.13	-0.05
Lying						0.00	0.00	-0.05	-0.02	0.01	0.04	-0.04	0.01	-0.03	-0.02
Rule following							0.13	0.01	0.03	0.14	0.06	0.06	0.05	-0.02	0.08
Altruism								0.52	0.17	0.28	0.26	0.40	0.32	0.23	0.26
Altruism Taking									0.12	0.22	0.21	0.27	0.28	0.22	0.24
Trust										0.22	0.34	0.42	0.32	0.09	0.17
Pos. Recipr. Low											0.55	0.24	0.31	0.18	0.23
Pos. Recipr. High												0.30	0.32	0.27	0.25
PGG Cont.													0.38	0.25	0.21
Reward Cons.- Int.														0.15	0.45
Reward Int.															0.38
Reward Cons.															

Table G.7: Biserial correlations between ends-versus-means and self-versus-other choice tasks

	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons. comp Int.	Reward Int.	Reward Cons.
SAL Trolley	-0.01	-0.04	0.03	-0.07	-0.04	0.01	-0.03	0.07	-0.03
Group Don.	0.06	0.04	-0.05	0.18	0.06	-0.03	0.11	-0.04	-0.02
Bribe	-0.05	-0.03	-0.16	-0.08	-0.21	-0.09	-0.08	-0.26	-0.25
Statement Choice	0.00	0.01	-0.15	0.05	-0.09	-0.01	0.07	-0.19	-0.05
Lying	-0.03	-0.05	-0.03	-0.03	0.00	-0.06	0.00	-0.03	-0.01
Rule following	0.15	0.08	0.04	0.14	0.13	0.04	0.15	-0.01	0.16

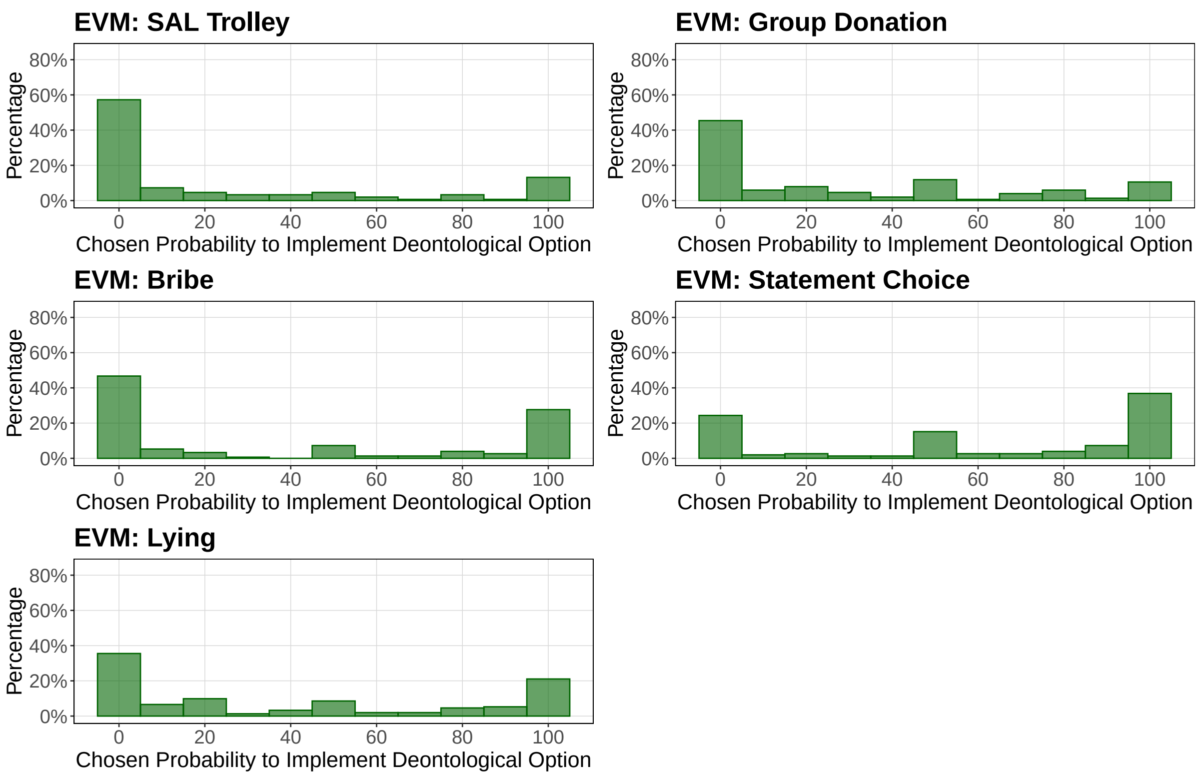
Table H.1: Correlation of the extensive margin of randomization across EVM choice tasks

	Does Rand. Group Don.	Does Rand. Bribe	Does Rand. Statement Choice	Does Rand. Lying
Does Rand. SAL Trolley	0.39	0.35	0.24	0.43
	[0.24, 0.52]	[0.21, 0.49]	[0.08, 0.38]	[0.30, 0.56]
Does Rand. Group. Don.		0.44	0.31	0.37
		[0.31, 0.56]	[0.15, 0.44]	[0.22, 0.50]
Does Rand. Bribe			0.33	0.31
			[0.18, 0.47]	[0.16, 0.45]
Does Rand. Statement Choice				0.31
				[0.16, 0.45]

Notes: The table displays Pearson correlation coefficients. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

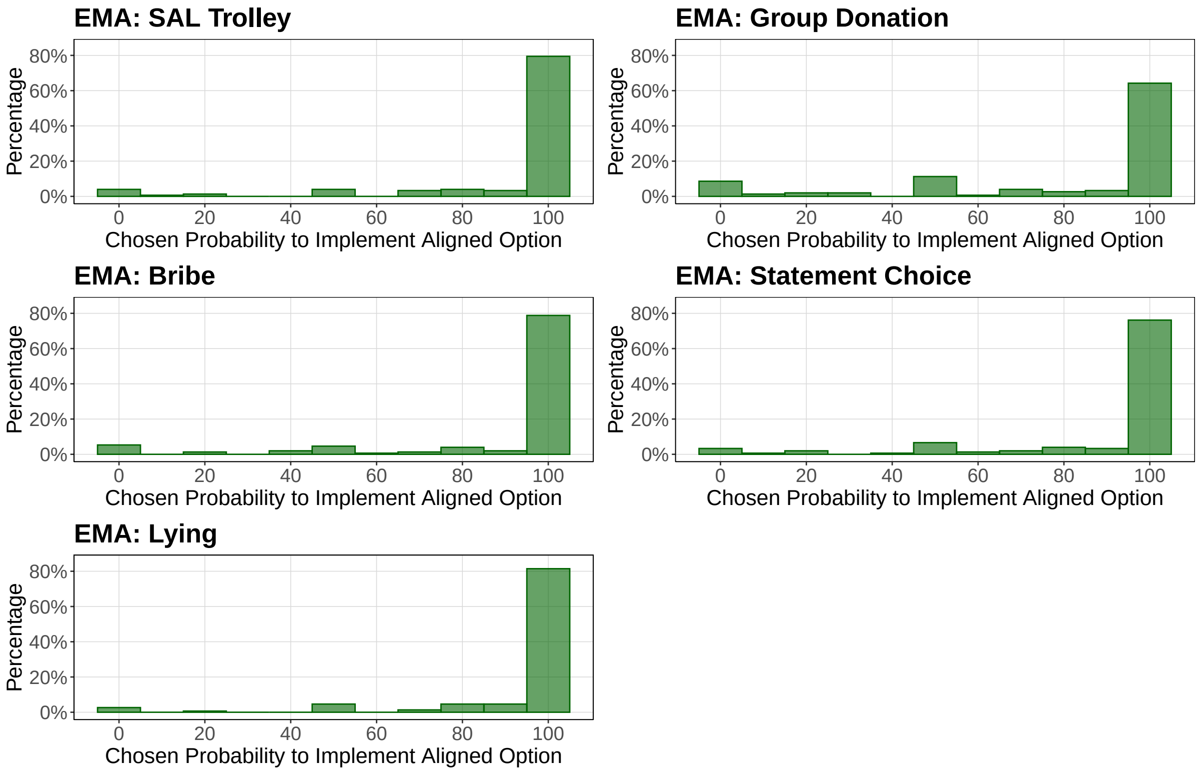
H Additional figures and tables for the randomization experiment

Figure H.1: Distribution of probabilities for the EVM choice tasks with randomization option



Notes: Each bar displays the fraction of subjects choosing the respective probability to implement the deontological option in the described decision-situation. See Section 3.1.2 for details on the situations and the definition of the options.

Figure H.2: Distribution of probabilities for the EMA choice tasks with randomization option



Notes: Each bar displays the fraction of subjects choosing the respective probability to implement the aligned option in the described decision-situation. See Section 3.1.2 for details on the situations and the definition of the options.

Table H.2: Correlation of the intensive margin of randomization across EVM choice tasks

	Group Don.	Bribe	Statement Choice	Lying
SAL Trolley	0.25	0.15	0.01	0.15
	[0.09, 0.39]	[-0.01, 0.30]	[-0.15, 0.17]	[-0.01, 0.30]
Group Don.		0.09	-0.03	0.09
		[-0.07, 0.25]	[-0.19, 0.13]	[-0.07, 0.25]
Bribe			0.05	0.12
			[-0.11, 0.21]	[-0.04, 0.27]
Statement Choice				0.02
				[-0.14, 0.18]

Notes: The table displays Pearson correlation coefficients. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

Figure H.3: Distribution of probabilities for the SVO choice tasks with randomization option

Notes: Each bar displays the fraction of subjects choosing the respective probability to implement the selfish option in the described decision-situation. See Section 3.2.2 for details on the situations and the definition of the options.

Table H.3: Correlation across EVM choice tasks using the tendency classification

	Group Don.	Bribe	Statement Choice	Lying
SAL Trolley	0.24	0.12	0.01	0.09
	[0.07, 0.40]	[-0.05, 0.29]	[-0.16, 0.19]	[-0.07, 0.26]
Group. Don.		0.12	-0.02	0.06
		[-0.06, 0.29]	[-0.20, 0.17]	[-0.12, 0.23]
Bribe			0.10	0.14
			[-0.08, 0.28]	[-0.04, 0.30]
Statement Choice				0.11
				[-0.07, 0.28]

Notes: The table displays Pearson correlation coefficients. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

Table H.4: Correlation of the extensive margin of randomization across SVO games

	Does Rand. Trust	Does Rand. Pos. Recipr. Low	Does Rand. Pos. Recipr. High	Does Rand. PGG Cont.
Does Rand. Altruism	0.25	0.20	0.30	0.29
	[0.08, 0.40]	[0.03, 0.35]	[0.14, 0.45]	[0.13, 0.44]
Does Rand. Trust		0.33	0.30	0.41
		[0.17, 0.47]	[0.15, 0.45]	[0.26, 0.54]
Does Rand. Pos. Recipr. Low			0.29	0.17
			[0.13, 0.43]	[0.01, 0.33]
Does Rand. Pos. Recipr. High				0.38
				[0.23, 0.51]

Notes: The table displays Pearson correlation coefficients. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

Table H.5: Correlation of the intensive margin of randomization across SVO games

	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.
Altruism	0.24	0.27	0.20	0.26
	[0.08, 0.39]	[0.11, 0.42]	[0.03, 0.35]	[0.10, 0.41]
Trust		0.39	0.17	0.45
		[0.24, 0.52]	[0.00, 0.32]	[0.30, 0.57]
Pos. Recipr. Low			0.25	0.46
			[0.09, 0.40]	[0.31, 0.58]
Pos. Recipr. High				0.21
				[0.04, 0.36]

Notes: The table displays Pearson correlation coefficients. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

Table H.6: Correlation across SVO games using the tendency classification

	Trust	Pos. Recripr. Low	Pos. Recripr. High	PGG Cont.
Altruism	0.25	0.31	0.22	0.27
	[0.07, 0.41]	[0.13, 0.47]	[0.04, 0.39]	[0.09, 0.44]
Trust		0.38	0.17	0.44
		[0.21, 0.53]	[-0.01, 0.34]	[0.28, 0.58]
Pos. Recripr. Low			0.29	0.45
			[0.12, 0.45]	[0.29, 0.59]
Pos. Recripr. High				0.24
				[0.07, 0.41]

Notes: The table displays Pearson correlation coefficients. Colors in heatmap change with each 0.1 correlation magnitude. Brackets display 95% confidence intervals.

I Research transparency

All experiments were preregistered at the AEA RCT Registry (#AEARCTR-0007714).. The preregistration includes details on the experimental design, the planned sample size, and an outline of the research question/hypotheses. In the following, we describe in more detail the mapping between the paper and preregistration.

I.1 Main experiment

Design and sample size. We implemented the experimental design as preregistered. We preregistered a target sample of 600 subjects completing both sessions. This sample size was chosen based on power calculations for the comparison between the *Hypothetical* and *Real* treatments of the SAL trolley. Assuming that 20% of subjects choose the deontological option in *Hypothetical* and a significance level of 5%, we have 80% power to detect a 10 pp. treatment effect (two-sample test of proportions) –that is, a reduction in the fraction of subjects choosing the deontological option to 10% or less, or an increase to 30% or more. In the experiment, 593 subjects completed both sessions. We subsequently applied the preregistered exclusion restrictions of (i) excluding the top 1% faster subjects and (ii) subjects who preferred giving 2 Euros to another participant over a 15 Euros donation to charity in the baseline task. This resulted in the final sample of 548 subjects.

Hypotheses. The preregistration specified the following three research questions: (i) Are there moral types that behave consistent with consequentialism and deontological ethics and can they be identified by observing choices in the trolley dilemma or other situations? (ii) Does this depend on whether decisions come with real consequences or are posed hypothetically? (iii) Is consequentialistic and deontologically motivated behavior correlated with general economic preferences (self-versus-other behavior) and other relevant dimensions (moral questionnaires)? We cover the existence and consistency of types (first research question) in Section 4. The second research question is covered in Section 4.1, paragraph "Robustness: Trolley treatments", and the third in Section 6.

I.2 Aligned experiment

Design and sample size. We implemented the experimental design as preregistered. We preregistered a target sample size of 120 subjects. The sample size was chosen based on power calculations for the comparison between the EVM tradeoff and EMA aligned choice tasks. Assuming that 30% of subjects choose the deontological option in each choice task and a significance level of 5%, we have 80% power to detect a 20 pp. treatment effect (two-sample test of proportions). In total, 135 subjects participated in the experiment. Applying the preregistered exclusion restrictions (same as in the main experiment) leaves us with a final sample of 122 subjects.

Hypotheses. The preregistration specified 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? Based on the research question, the preregistered hypothesis was that a large majority of subjects would choose the aligned option in each EMA choice task. The results are covered in Section 4, paragraph "Robustness: Aligned treatment."

I.3 Randomization experiment

Design and sample size. We implemented the experimental design as preregistered. We preregistered a target sample size of 480 subjects. The sample size was chosen based on power calculations to test whether the fraction of subjects randomizing is significantly different from zero. Assuming that 30% of subjects randomize in the choice tasks and a significance level of 5%, we have 80% power to detect a 10 pp. fraction of subjects randomizing (one-sample test of proportions). In total, 478 subjects donated. We again preregistered and applied the same exclusion restrictions as before, leaving us with a final sample of 443 subjects.

Hypotheses. The preregistration specified the following research question: Is the lack of consistency among EVM decisions in the main experiment a consequence of subjects struggling with the difficulty of choosing in EVM decisions? Based on this research question, we preregistered the following three hypotheses: (i) For each decision situation in the EVM tradeoff decisions treatment, the fraction randomizing is significantly different from zero. (ii) The average fraction randomizing in the EVM decisions treatment is significantly higher than the average fraction randomizing in the EMA decisions. (iii) The average fraction randomizing in the EVM tradeoff decisions treatment is significantly higher than the average fraction randomizing in the SVO decisions. The results testing the first two hypotheses are covered in Section 4.3. The results for the third hypothesis are covered in Section 5.2.

J Instructions

Contents

J.1 General information . . . 69
J.2 Ends-versus-means block . . . 71
J.2.1 Trolley dilemma - Tradeoff version . . . 71
J.2.2 Trolley dilemma – aligned version . . . 77
J.2.3 Trolley dilemma – Randomization version . . . 84
J.2.4 Baseline choice . . . 84
J.2.5 Lying game - Tradeoff version . . . 85
J.2.6 Lying game - aligned version . . . 86
J.2.7 Lying game - Randomization version . . . 88
J.2.8 Bribe game - Tradeoff version . . . 89
J.2.9 Bribe game - aligned version . . . 90
J.2.10 Bribe game - Randomization version . . . 91
J.2.11 Group decision game - Tradeoff version . . . 92
J.2.12 Group decision game - aligned version . . . 93
J.2.13 Group decision game - Randomization version . . . 95
J.2.14 Statement choice - Tradeoff version . . . 96
J.2.15 Statement choice - aligned version . . . 97
J.2.16 Statement choice - Randomization version . . . 99
J.2.17 Rule following game - Tradeoff version . . . 100
J.2.18 Rule following game - aligned version . . . 100
J.3 Self-versus-others block . . . 100
J.3.1 Dictator game giving - Main version . . . 100
J.3.2 Dictator game giving - Randomization version . . . 101
J.3.3 Dictator game taking - Main version . . . 102
J.3.4 Trust game sender - Main version . . . 102
J.3.5 Trust game sender - Randomization version . . . 103
J.3.6 Trust game receiver - Main version . . . 105
J.3.7 Trust game receiver - Randomization version . . . 105
J.3.8 Public goods game - Main version . . . 107
J.3.9 Public goods game - Randomization version . . . 108
J.3.10 Moral luck game . . . 110

J.1 General information

Welcome to the study

Welcome, and thank you for your interest in today's online study! Please note that you may only participate in this study once. Also, you may only participate if you have registered for this study in our participation database.

As already announced during registration, this study consists of two parts. You will complete the second part exactly in one week at the same time. For completing both parts in full, you will receive a lump sum of 12 EUR upon completion of the study.

You will make decisions on the computer in this study. Your decisions will allow you to earn extra money. All payments, i.e., both the compensation for your participation and any additional payments based on your decisions, will be sent to you by bank transfer following the second part.

Please note: You will receive the payment for your participation of 12 EUR as well as any additional money based on your decisions only if you successfully complete both parts. Therefore, please make sure to show up for the second part next week as well and complete it entirely.

Structure of the study and your payoff

In the following, you will make different decisions in each of several successive sections. The decisions in each section are completely independent of each other. This means your decisions in one section do not affect the consequences or possible payoffs of another section. Furthermore, similar-sounding decision situations in different sections do not mean that your decisions should necessarily be similar as well. You will see the number of the current section in the upper right corner of each section.

From all the decisions you will make today, a computer will randomly select one decision. Each of the decisions has the same chance of being selected. This selection is independent of your decisions.

The consequence of the selected decision will then be implemented exactly as described in the decision. Since one of your decisions will actually be implemented, you should consider each decision carefully and treat each decision as if it were actually implemented.

In what follows, this is always implicitly assumed. For example, suppose a decision is described as giving you an amount of money as a consequence. In that case, you will receive that amount exactly as described if that decision is indeed randomly selected.

Donation to the *Förderkreis*

Some decisions in today's part are about the possibility of making a donation to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* (in short: the *Förderkreis*). The *Förderkreis* is a regional aid organization from Bonn.

Information about the Förderkreis

The *Förderkreis* comprehensively supports young people with cancer and their families in dealing with the disease. The organization offers psychological support, organizes leisure activities as well as aftercare, and offers school support. The projects and participations of the *Förderkreis* at a glance are:

- Team Bärenstark: Practical help during the treatment period for patients and relatives.
- Psychological and psychosocial counseling for patients and relatives
- Pedagogical support in the hospital
- Financing of hospital clowns and music therapy
- Follow-up care by the KoCkPiT team (Children's Palliative Care Team Bonn)
- Support for parents of sick children and adolescents
- Support for orphaned parents

So if in the following, a donation to the *Förderkreis* is mentioned, it is always a donation to the organization described above, with which the listed projects are supported.

If in the following a donation is initiated by your decisions, it will always be paid by the scientists involved from study funds. So you do not have to donate (or pay) any amount of money personally. If you decide against a donation, the amount will not be donated accordingly.

Consequences for other study participants

In some sections, your decisions have consequences for other study participants. For example, in some choices, you can send money to a randomly selected other person who is also participating in this study.

Important: The other people will not learn about your decisions or personal information about you at any time. Notably, the other individuals will also receive payouts through other situations, which are random in amount. This ensures that the other people cannot draw any conclusions about your decisions from the amount of the payoffs at the end of the trial.

On the next page, the individual sections begin.

J.2 Ends-versus-means block

J.2.1 Trolley dilemma - Tradeoff version

The consequences of the next decision

[Treatment *Hypothetical*]

For the next section of this study, the decisions are hypothetical in nature. This means that the consequence of your decision in the next section will not be implemented in real terms. Thus, the following section is purely a thought experiment.

[Treatment *Real*]

For your decision in the next section, there is a probability of 1 in 10, i.e., 10%, that it will be implemented as described. At the end of the study, a computer-generated random number will determine whether your decision will actually be implemented. If so, the consequences of your decision will be realized exactly as described. Therefore, since your decision may actually be implemented, you should think carefully about the decision.

Information

Below you will read important information that is relevant to your decision later.
After you have made your decision, we will ask you four simple questions about the information and instructions presented bellow. For each question you then answer correctly, you will receive an additional 0.5 EUR.
Please read all the information carefully. Not only is the information important for your decision, but you will also potentially receive a higher payout.

Information about tuberculosis

What is tuberculosis?
Tuberculosis – also called Phthisis or White Death – is an infectious disease, which is caused by bacteria. Roughly one-third of all humans are infected with the pathogen of tuberculosis. Active tuberculosis breaks out among 5 to 10% of all those infected. Tuberculosis is primarily airborne. This is also why quick treatment is necessary.

What are the symptoms and consequences of tuberculosis?
Tuberculosis patients often suffer from very unspecific symptoms like fatigue, the feeling of weakness, lack of appetite, and weight loss. At an advanced stage of lung tuberculosis, the patient coughs up blood, leading to the so-called rush of blood. Without treatment, a person with tuberculosis dies with a probability of 43%.

How prevalent is tuberculosis?

In the year 2018, about 10 million people have been recorded as falling ill with active tuberculosis. Almost 1.5 million people die of tuberculosis each year. This means that tuberculosis causes more annual deaths than HIV or malaria.

Is tuberculosis curable?
According to the World Health Organization (WHO), the United Nations agency for international public health, "tuberculosis is preventable and curable". Treatment takes place by taking antibiotics several times a week over a period of 6 months. It is important to take the medication consistently. Since 2000, an estimated 53 million lives have been saved through effective diagnosis and treatment of tuberculosis.

The success rate of treatment for a new infection is usually over 85%.
The preceding figures and information were provided by WHO and are freely available. For more details click here (link opens in new tab).

Operation ASHA

Operation ASHA is a charity organization specialized since 2005 on treating tuberculosis in disadvantaged communities. The work of *Operation ASHA* is based on the insight that the biggest obstacle for the treatment of tuberculosis is the interruption of the necessary 6-month-long regular intake of medication.

For a successful treatment, the patient has to come to a medical facility twice a week – more than 60 times in total – to take the medication. Interruption or termination of the treatment is fatal because this strongly enhances the development of a drug-resistant form of tuberculosis. This form of tuberculosis is much more difficult to treat and almost always leads to death.

The concept of *Operation ASHA*
To overcome this problem, *Operation ASHA* developed a concept that guarantees regular treatment through immediate spatial proximity to the patient. A possible non-adherence is additionally prevented by visiting the patient at home.

By now, *Operation ASHA* runs more than 360 treatment centers, almost all of which are located in the poorer regions of India. More than 60,000 sick persons have been identified and treated that way.

Operation ASHA is an internationally recognized organization, and its success has been covered by the New York Times, BBC, and Deutsche Welle, for example. The MIT and the University College London have already conducted research projects about the fight against tuberculosis in cooperation with *Operation ASHA*. The treatment method employed by *Operation ASHA* is described by the World Health Organization (WHO) as "highly efficient and cost-effective".

The impact of a donation to *Operation ASHA*

By donating money to *Operation ASHA*, it is possible to save people from death by tuberculosis.

To save a person's life means here to successfully cure a person with tuberculosis, who otherwise would die because of the tuberculosis. A donation of 380 EUR ensures that at least one human life can be expected to be saved. The information used to calculate the donation amount is obtained from public statements from the World Health Organization (WHO), peer-reviewed research studies, Indian Government statistics, and published figures from *Operation ASHA*.

In the calculation, information was conservatively interpreted, or a pessimistic number was used so that the donation amount of 380 EUR is, in the case of doubt, higher than the actual costs to save a human life. In addition, in the calculation of the treatment success rate of *Operation ASHA*, the mortality rate for alternative treatment by the state tuberculosis program in India and the different detection rates for new cases of tuberculosis are included.

In the context of this study, an agreement made with *Operation ASHA* will ensure that 100% of the donation will be used exclusively for the diagnosis and treatment of tuberculosis patients. This means that every Euro of the donation amount goes directly to saving human lives, and no other costs will be covered. Based on a very high number of cases, the contribution of a donation of 380 EUR can be summarized as follows:

With a donation of 380 EUR 5 additional patients infected with tuberculosis can be treated through *Operation ASHA*.

If these 5 persons are not treated through *Operation ASHA*, it is expected that one patient will die.

If, through the donation of 380 EUR all 5 patients are treated, it is expected that no patient will die.

Based on these calculations it can be concluded that the donation of 380 EUR will save one human life. Going further, this also means that two lives are saved by a donation of 760 EUR, and three lives are saved by a donation of 1140 EUR.

The relationship of a donation of 380 EUR to Operation ASHA and a human life is illustrated in the following graph: [Figure here]

So, in summary, the following can be said about the effect of a donation for a given number of 5 patients suffering from tuberculosis:

- Donation results in saving one person from death.
- Failure to donate results in the death of one person.

Whenever the life or death of a person is mentioned in connection with a donation, the above-mentioned facts are implicitly meant. The same applies when the life or death of several people is involved.

The concrete situation

As mentioned before, *Operation ASHA* operates numerous treatment centers in India. In particular, *Operation ASHA* operates in the two Indian states of [State A] and [State B].

[State A]

In this experiment, 5 people were identified in [State A] who are infected by tuberculosis. For these people, the donation amount required to save a life was initiated to *Operation ASHA*. Through this, only the people who are sick in [State A] will be treated. This will result in a person from [State A] being saved from death by tuberculosis.

If the initiated donation amount were not realized, a person in [State A] would die from tuberculosis.

[State B]

After the donation was initiated, 15 people suffering from tuberculosis were identified in [State B]. Without donation and subsequent treatment by *Operation ASHA*, three of the people from [State B] who became ill would die from tuberculosis.

In the event that the people in [State B] are helped through the donation to *Operation ASHA* necessary to save three lives, these three people would be saved from death by tuberculosis.

Redirecting the donation

The initiated donation to patients in [State A] can be redirected to help save patients in [State B] from death by tuberculosis. The amount would be multiplied to save three lives in [State B].

Summary

Tuberculosis is a worldwide common bacterial infectious disease. The success rate of medical treatment of a new disease is very high. Nevertheless, close to 1.5 million people die every year from tuberculosis. The biggest obstacle to the curing of tuberculosis is the potential stopping of continuous treatment with antibiotics. The concept of *Operation ASHA* is therefore based on the immediate proximity to the patient as well as the control and recording of the regular intake of medication. A donation of 380 EUR to Operation ASHA can save a human life. A donation of 760 EUR can save two lives, and a donation of 1140 EUR can save three lives.

How is the donation connected to the saving of a life?

The donation of 380 EUR already accounts for the fact that someone inflicted with the illness could have survived without treatment by *Operation ASHA*; i.e., instead of through *Operation ASHA*, they could have received treatment through other actors (such as the public health system). The amount is, therefore, sufficient for the diagnosis and complete treatment of multiple sufferers.

What does it mean to "save a life"?
To save a life means here the successful curing of a person suffering from tuberculosis, who otherwise would die because of the Tuberculosis. This means: The amount donated is sufficient to identify and treat enough people with tuberculosis that there is at least one person among them who would otherwise have died from tuberculosis in expectation. To summarize:

- A donation has the effect of saving a person from death.
- Failure to donate results in the death of a person.

The initial situation
In [State A], 5 people suffering from tuberculosis were identified. For these, the necessary donation amount was initiated to save one person from death due to tuberculosis. Thus, without further intervention, none of these diseased people will die of tuberculosis.

After initiating this donation, an additional 15 people suffering from tuberculosis were identified in [State B]. Without further intervention, three of these diseased people would die of tuberculosis. If the initiated donation amount is redirected, it will be multiplied to save three people in [State B] from dying of tuberculosis.

Your decision

So, in terms of the context just described, the situation is as follows:

Without intervention, three people in [State B] will die of tuberculosis. By intervening, these three people can be saved. However, as a direct consequence of this intervention, one person in [State A] will die of tuberculosis.

You will have the opportunity to choose between intervening and not intervening in this situation:

Do not intervene
Do not intervene: do not divert the donation.
Consequence: You decide not to intervene. This means that you decide against an intervention on your end that would cause a patient in [State A] to die from tuberculosis. In return, three people from [State B] will not be saved from dying from tuberculosis.

Intervene
Intervene: redirect the donation.

Consequence: You decide to intervene. This means that you decide in favor of an intervention that causes a patient in [State A] to die from tuberculosis. In return, your redirected donation will save three people from [State B] from dying of tuberculosis.

In a moment, you will have the opportunity to make this decision using an animation. The next page will explain how this animation works.

The decision animation

In a moment, you will have the opportunity to make your decision with the help of an animation. The initial situation of the animation is shown in the following figure:

[animation figure]
The animation depicts the decision situation just described. As soon as the animation starts, the skull shown moves slowly to the right towards the three people. In case you don't see three people on the right or they are outside the white border, please click [here].

You can change the path of the skull using the slider located to the left of the junction (blue line). You do this by dragging the slider with your computer mouse all the way down and then releasing it. This is necessary because otherwise, the slider will move back to its original position. If the slider is dragged all the way down and released, the downward junction will open so that the skull on the track will move down to the single person.

You can make your decision before the skull reaches the dashed gray line in front of the junction. After arriving at the line, moving the slider will have no effect.

Hence, you have two choices, with the consequences described earlier:

Do not intervene
Do not intervene: do not move the slider, the junction will not open.
Consequence: You decide not to intervene. This means that you decide against an intervention on your end that would cause a patient in [State A] to die from tuberculosis. In return, three people from [State B] will not be saved from dying from tuberculosis.

Intervene
Intervene: move the slider, the junction will open.
Consequence: You decide to intervene. This means that you decide in favor of an intervention that causes a patient in [State A] to die from tuberculosis. In return, your redirected donation will save three people from [State B] from dying of tuberculosis.

On the following screen, you will have the opportunity to familiarize yourself with the controller. You will be able to make your actual decision afterward.

Testing the slider

You will now have the opportunity to try out the slider of the animation. Note that you need to drag the slider all the way down and then release it to open the branch.

[slider test]

Click “Next” when you are sufficiently familiar with the slider.

On the next screen you can now make your decision. Note that the animation will start directly.

As a reminder:

[Treatment *Hypothetical*]
As previously described, this is a hypothetical decision that is not actually implemented.

[Treatment *Real*]
As previously described, your decision may actually be implemented, resulting in actual consequences.

[Trolley decision]

J.2.2 Trolley dilemma – aligned version

The consequences of the next decision

[Treatment *Hypothetical*]

For the next section of this study, the decisions are hypothetical in nature. This means that the consequence of your decision in the next section will not be implemented in real terms. Thus, the following section is purely a thought experiment.

[Treatment *Real*]

For your decision in the next section, there is a probability of 1 in 10, i.e., 10%, that it will be implemented as described. At the end of the study, a computer-generated random number will determine whether your decision will actually be implemented. If so, the consequences of your decision will be realized exactly as described. Therefore, since your decision may actually be implemented, you should think carefully about the decision.

Information

Below you will read important information that is relevant to your decision later.

After you have made your decision, we will ask you four simple questions about the information and instructions presented bellow. For each question you then answer correctly, you will receive an additional 0.5 EUR.
Please read all the information carefully. Not only is the information important for your decision, but you will also potentially receive a higher payout.

Information about tuberculosis

What is tuberculosis?
Tuberculosis – also called Phthisis or White Death – is an infectious disease, which is caused by bacteria. Roughly one-third of all humans are infected with the pathogen of tuberculosis. Active tuberculosis breaks out among 5 to 10% of all those infected. Tuberculosis is primarily airborne. This is also why quick treatment is necessary.

What are the symptoms and consequences of tuberculosis?
Tuberculosis patients often suffer from very unspecific symptoms like fatigue, the feeling of weakness, lack of appetite, and weight loss. At an advanced stage of lung tuberculosis, the patient coughs up blood, leading to the so-called rush of blood. Without treatment, a person with tuberculosis dies with a probability of 43%.

How prevalent is tuberculosis?
In the year 2018, about 10 million people have been recorded as falling ill with active tuberculosis. Almost 1.5 million people die of tuberculosis each year. This means that tuberculosis causes more annual deaths than HIV or malaria.

Is tuberculosis curable?
According to the World Health Organization (WHO), the United Nations agency for international public health, "tuberculosis is preventable and curable". Treatment takes place by taking antibiotics several times a week over a period of 6 months. It is important to take the medication consistently. Since 2000, an estimated 53 million lives have been saved through effective diagnosis and treatment of tuberculosis.

The success rate of treatment for a new infection is usually over 85%.
The preceding figures and information were provided by WHO and are freely available. For more details click here (link opens in new tab).

Operation ASHA

Operation ASHA is a charity organization specialized since 2005 on treating tuberculosis in disadvantaged communities. The work of *Operation ASHA* is based on the insight that the biggest obstacle for the treatment of tuberculosis is the interruption of the necessary 6-month-long regular intake of medication.

For a successful treatment, the patient has to come to a medical facility twice a week – more than 60 times in total – to take the medication. Interruption or termination of the treatment is fatal because this strongly enhances the development of a drug-resistant form of tuberculosis. This form of tuberculosis is much more difficult to treat and almost always leads to death.

The concept of *Operation ASHA*

To overcome this problem, *Operation ASHA* developed a concept that guarantees regular treatment through immediate spatial proximity to the patient. A possible non-adherence is additionally prevented by visiting the patient at home.

By now, *Operation ASHA* runs more than 360 treatment centers, almost all of which are located in the poorer regions of India. More than 60,000 sick persons have been identified and treated that way.

Operation ASHA is an internationally recognized organization, and its success has been covered by the New York Times, BBC, and Deutsche Welle, for example. The MIT and the University College London have already conducted research projects about the fight against tuberculosis in cooperation with *Operation ASHA*. The treatment method employed by *Operation ASHA* is described by the World Health Organization (WHO) as “highly efficient and cost-effective”.

The impact of a donation to *Operation ASHA*

By donating money to *Operation ASHA*, it is possible to save people from death by tuberculosis.

To save a person’s life means here to successfully cure a person with tuberculosis, who otherwise would die because of the tuberculosis. A donation of 380 EUR ensures that at least one human life can be expected to be saved. The information used to calculate the donation amount is obtained from public statements from the World Health Organization (WHO), peer-reviewed research studies, Indian Government statistics, and published figures from *Operation ASHA*.

In the calculation, information was conservatively interpreted, or a pessimistic number was used so that the donation amount of 380 EUR is in the case of doubt higher than the actual costs to save a human life. In addition, in the calculation of the treatment success rate of *Operation ASHA*, the mortality rate for alternative treatment by the state tuberculosis program in India and the different detection rates for new cases of tuberculosis are included.

In the context of this study, an agreement made with *Operation ASHA* will ensure that 100% of the donation will be used exclusively for the diagnosis and treatment of tuberculosis patients. This means that every Euro of the donation amount goes directly to saving human lives, and no other costs will be covered. Based on a very high number of cases, the contribution of a donation of 380 EUR can be summarized as follows:

With a donation of 380 EUR 5 additional patients infected with tuberculosis can be treated through *Operation ASHA*.

If these 5 persons are not treated through *Operation ASHA*, it is expected that one patient will die.

If, through the donation of 380 EUR all 5 patients are treated, it is expected that no patient will die.

Based on these calculations it can be concluded that the donation of 380 EUR will save one human life. Going further, this also means that two lives are saved by a donation of 760 EUR, and three lives are saved by a donation of 1140 EUR.

The relationship of a donation of 380 EUR to Operation ASHA and a human life is illustrated in the following graph: [Figure here]

So, in summary, the following can be said about the effect of a donation for a given number of 5 patients suffering from tuberculosis:

- Donation results in saving one person from death.
- Failure to donate results in the death of one person.

Whenever the life or death of a person is mentioned in connection with a donation, the above-mentioned facts are implicitly meant. The same applies when the life or death of several people is involved.

The concrete situation

As mentioned before, *Operation ASHA* operates numerous treatment centers in India. In particular, *Operation ASHA* operates in the two Indian states of [State A] and [State B].

[State A]
In this experiment, 15 people were identified in [State A] who are infected by tuberculosis. For these people, the donation amount required to save a life was initiated to *Operation ASHA*. Through this, only the people who are sick in [State A] will be treated. This will result in three people from [State A] being saved from death by tuberculosis.

If the initiated donation amount were not realized, three people in [State A] would die from tuberculosis.

[State B]
After the donation was initiated, 5 people suffering from tuberculosis were identified in [State B].

Without donation and subsequent treatment by *Operation ASHA*, one person from [State B] who became ill would die from tuberculosis.

In the event that the people in [State B] are helped through the donation to *Operation ASHA* necessary to save one live, these one person would be saved from death by tuberculosis.

Redirecting the donation

The initiated donation to patients in [State A] can be redirected to help save patients in [State B] from death by tuberculosis. The amount would be reduced to save one live in [State B].

Summary

Tuberculosis is a worldwide common bacterial infectious disease. The success rate of medical treatment of a new disease is very high. Nevertheless, close to 1.5 million people die every year from tuberculosis. The biggest obstacle to the curing of tuberculosis is the potential stopping of continuous treatment with antibiotics. The concept of *Operation ASHA* is therefore based on the immediate proximity to the patient as well as the control and recording of the regular intake of medication. A donation of 380 EUR to Operation ASHA can save a human life. A donation of 760 EUR can save two lives, and a donation of 1140 EUR can save three lives.

How is the donation connected to the saving of a life?
The donation of 380 EUR already accounts for the fact that someone inflicted with the illness could have survived without treatment by *Operation ASHA*; i.e., instead of through *Operation ASHA*, they could have received treatment through other actors (such as the public health system). The amount is, therefore, sufficient for the diagnosis and complete treatment of multiple sufferers.

What does it mean to "save a life"?
To save a life means here the successful curing of a person suffering from tuberculosis, who otherwise would die because of the Tuberculosis. This means: The amount donated is sufficient to identify and treat enough people with tuberculosis that there is at least one person among them who would otherwise have died from tuberculosis in expectation. To summarize:

- A donation has the effect of saving a person from death.
- Failure to donate results in the death of a person.

The initial situation
In [State A], 15 people suffering from tuberculosis were identified. For these, the necessary donation amount was initiated to save three people from death due to tuberculosis. Thus, without further intervention, none of these diseased people will die of tuberculosis.

After initiating this donation, an additional 5 people suffering from tuberculosis were identified in [State B]. Without further intervention, one of these diseased people would die of tuberculosis. If the initiated donation amount is redirected, it will be reduced to save one person in [State B] from dying of tuberculosis.

Your decision

So, in terms of the context just described, the situation is as follows:

Without intervention, one person in [State B] will die of tuberculosis. By intervening, this person can be saved. However, as a direct consequence of this intervention, three people in [State A] will die of tuberculosis.

You will have the opportunity to choose between intervening and not intervening in this situation:

Do not intervene
Do not intervene: do not divert the donation.
Consequence: You decide not to intervene. This means that you decide against an intervention on your end that would cause three patients in [State A] to die from tuberculosis. In return, one person from [State B] will not be saved from dying from tuberculosis.

Intervene
Intervene: redirect the donation.
Consequence: You decide to intervene. This means that you decide in favor of an intervention that causes three patients in [State A] to die from tuberculosis. In return, your redirected donation will save one person from [State B] from dying of tuberculosis.

In a moment, you will have the opportunity to make this decision using an animation. The next page will explain how this animation works.

The decision animation

In a moment, you will have the opportunity to make your decision with the help of an animation. The initial situation of the animation is shown in the following figure:

[animation figure]
The animation depicts the decision situation just described. As soon as the animation starts, the skull shown moves slowly to the right towards the three people. In case you don't see three people on the right or they are outside the white border, please click [here].

You can change the path of the skull using the slider located to the left of the junction (blue line). You do this by dragging the slider with your computer mouse all the way down and then releasing it. This is necessary because otherwise, the slider will move back to its original position. If the slider is dragged all the way down and released, the downward junction will open so that the skull on the track will move down to the single person.

You can make your decision before the skull reaches the dashed gray line in front of the junction. After arriving at the line, moving the slider will have no effect.

Hence, you have two choices, with the consequences described earlier:

Do not intervene
Do not intervene: do not move the slider, the junction will not open.
Consequence: You decide not to intervene. This means that you decide against an intervention on your end that would cause three patients in [State A] to die from tuberculosis. In return, one person from [State B] will not be saved from dying from tuberculosis.

Intervene
Intervene: move the slider, the junction will open.
Consequence: You decide to intervene. This means that you decide in favor of an intervention that causes three patients in [State A] to die from tuberculosis. In return, your redirected donation will save one person from [State B] from dying of tuberculosis.

On the following screen, you will have the opportunity to familiarize yourself with the controller. You will be able to make your actual decision afterward.

Testing the slider

You will now have the opportunity to try out the slider of the animation. Note that you need to drag the slider all the way down and then release it to open the branch.

[slider test]

Click "Next" when you are sufficiently familiar with the slider.

On the next screen you can now make your decision. Note that the animation will start directly.

As a reminder:

[Treatment *Hypothetical*]
As previously described, this is a hypothetical decision that is not actually implemented.

[Treatment *Real*]
As previously described, your decision may actually be implemented, resulting in actual consequences.

[Trolley decision]

J.2.3 Trolley dilemma – Randomization version

[Instructions before the ”The decision animation” part are identical to the tradeoff/aligned version]

Selection of Options

You choose between the two options by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Option A (do not redirect the donation) will be implemented with a probability of 0%, and Option B (redirect the donation) with a probability of 100%.
- If you enter 1, Option A will be implemented with a probability of 10%, and Option B with a probability of 90%.
- If you enter 2, Option A will be implemented with a probability of 20%, and Option B with a probability of 80%.
- And so on, up to the number 10, at which point Option A will be implemented with a probability of 100% and Option B with a probability of 0%.

Decision

Choose a number from 0 to 10:

Consequence of Your Decision

The donation will not be redirected (Option A) with a probability of *–Enter a number–* %.

The donation will be redirected (Option B) with a probability of *–Enter a number–* %.

J.2.4 Baseline choice

In this section, you have a choice between two options, Option A and Option B. The two options differ in whether a sum of money is donated to a charity or another study participant receives a sum of money. The two options are as follows:

Option A

If you choose option A, 15 EUR will be donated to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.*

Option B

If you choose option B, another study participant will receive 2 EUR, and no money will be donated to the *Förderkreis*. The other study participant will not receive any additional information about your decision.

J.2.5 Lying game - Tradeoff version

In this section, you have been paired with another person. This person, like you, is a participant in this study.

Options

In this section, there are two possible payout options that you and the other person can achieve at the end of the study. The two options are:

Option A
15 EUR will be donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V.

Option B
The other person receives 2 EUR and no money is donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V.

The other person can choose between these two options, that is, between option A or option B. **However, the other person does not know what the exact consequences of the options are. In particular, the other person does not know which option leads to which payout.** Only you know what the consequences of the options are. The only information that the other person will have prior to their decision is a single message that you can send to the person.

Your decision in this section

You have a choice between the following two messages that you can send to the other person:

Message 1
"Option A gives you a higher personal cash payment than Option B".

Message 2
"Option B gives you a higher personal cash payment than Option A".

The sent message will be displayed to the other person exactly as shown above. In addition, the person will not learn which payoffs were associated with which option (also not afterward). Since the person will receive additional payments from other tasks in this study, as described at the beginning of the study, the person will also not learn whether your message was right or wrong.

The person was told that the message received could be true or false and that he or she would not receive any information about what the consequences of the options were. In such a situation, the other person follows the action prescribed in the message in almost all cases (>90%). Thus, you can assume with a very high probability the following:

- If you send message 1, the other person will choose option A (without knowing the exact consequences).
- If you send message 2, the other person will choose option B (without knowing the exact consequences).

Hence, the sequence of events is the following: You decide which of the two messages you want to send. Based on that, the other person will choose the respective option. The other person's decision will then determine the corresponding payments that will be made at the end of the study.

Summary

In summary, by choosing the message, you can have an impact on the payments and donations that are triggered. The other person is aware that (1) the message may be right or wrong, (2) they will not know which payoffs were associated with which option and whether the message was right or wrong.

[Comprehension questions]

Decision screen

You can now choose between the two messages. Click here if you would like to read the instructions again. As a reminder, here are the two messages:

- **Message 1:** "Option A gives you a higher personal cash payment than Option B."
- **Message 2:** "Option B gives you a higher personal cash payment than Option A."

Your choice: [ ]

J.2.6 Lying game - aligned version

In this section, you have been paired with another person. This person, like you, is a participant in this study.

Options

In this section, there are two possible payout options that you and the other person can achieve at the end of the study. The two options are:

Option A
15 EUR will be donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V.

Option B
The other person receives 2 EUR and no money is donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V.

The other person can choose between these two options, that is, between option A or option B. **However, the other person does not know what the exact consequences of the options are. In particular, the other person does not know which option leads to which payout.** Only you know what the consequences of the options are. The only information that the other person will have prior to their decision is a single message that you can send to the person.

Your decision in this section

You have a choice between the following two messages that you can send to the other person:

Message 1
"Option A triggers a higher donation to a charity than Option B".

Message 2
"Option B triggers a higher donation to a charity than Option A".

The sent message will be displayed to the other person exactly as shown above. In addition, the person will not learn which payoffs were associated with which option (also not afterward). Since the person will receive additional payments from other tasks in this study, as described at the beginning of the study, the person will also not learn whether your message was right or wrong.

The person was told that the message received could be true or false and that he or she would not receive any information about what the consequences of the options were. In such a situation, the other person follows the action prescribed in the message in almost all cases (>90%). Thus, you can assume with a very high probability the following:

- If you send message 1, the other person will choose option A (without knowing the exact consequences).
- If you send message 2, the other person will choose option B (without knowing the exact consequences).

Hence, the sequence of events is the following: You decide which of the two messages you want to send. Based on that, the other person will choose the respective option. The other person's decision will then determine the corresponding payments that will be made at the end of the study.

Summary

In summary, by choosing the message, you can have an impact on the payments and donations that are triggered. The other person is aware that (1) the message may be right or wrong, (2) they will not know which payoffs were associated with which option and whether the message was right or wrong.

[Comprehension questions]

Decision screen

You can now choose between the two messages. Click here if you would like to read the instructions again. As a reminder, here are the two messages:

- **Message 1:** "Option A triggers a higher donation to a charity than Option B."
- **Message 2:** "Option B triggers a higher donation to a charity than Option A."

Your choice: []

J.2.7 Lying game - Randomization version

[Instructions before the decision screen are identical to the tradeoff/aligned version]

Decision screen

You can now choose between the two messages. Click here if you would like to read the instructions again. As a reminder, here are the two messages:

- **Message 1:** "Option A gives you a higher personal cash payment than Option B."
- **Message 2:** "Option B gives you a higher personal cash payment than Option A."

You choose between the two messages by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Message 1 will be sent with a probability of 0% and Message 2 with a probability of 100%.
- If you enter 1, Message 1 will be sent with a probability of 10% and Message 2 with a probability of 90%.
- If you enter 2, Message 1 will be sent with a probability of 20% and Message 2 with a probability of 80%.
- And so on, up to the number 10, at which point Message 1 will be sent with a probability of 100% and Message 2 with a probability of 0%.

Decision

Choose a number from 0 to 10: []

Consequence of Your Decision

Message 1 will be sent with a probability of *–Enter a number–* %.

Message 2 will be sent with a probability of *–Enter a number–* %.

J.2.8 Bribe game - Tradeoff version

In this section, you have been matched with another person. Like you, this person is a participant in the current study. In this section, you have a choice between two options, Option A and Option B. The two options are as follows:

Option A
If you choose Option A, you will help children with cancer. Your choice of Option A will trigger a donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.*
Option B
If you choose option B, you will send the other person 2 EUR. However, in this case, the donation otherwise triggered by option A will not be given to *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.*

Additional consequence

The other person also has the choice between two options, namely between option C and option D. If the other person chooses option C, an additional 20 EUR will be donated to the *Förderkreis* If the person chooses option D, the person will receive 2 EUR for themselves as an additional payout.

The person has chosen to make a choice between the two options, C and D, dependent on **your decision** between A and B. This means the other person has **obligatorily** specified the following: The person will choose option D and thus the additional payout for himself. The choice of option D can only be prevented if you choose option B and send the other person 2 EUR. In this and only in this case, the other person will choose option C and thus trigger the donation. Thus, the consequences are as follows:

- You choose option A ⇒ The other person chooses option D.
- You choose option B ⇒ The other person chooses option C.

In other words, if you choose option A and thus trigger a donation of 15 EUR, the other person will stick with option D and receive 2 EUR. If you choose to send the 2 EUR to the other person by choosing option B, the other person will choose option C and choose the donation of 20 EUR.

Decision screen

You can now choose between the two options. Click here if you would like to read the instructions again. As a reminder, here are the two options:

- **Option A:** I help children with cancer by donating 15 EUR to the *Förderkreis*.
- **Option B:** I send 2 EUR to the other person.

Your choice: []

J.2.9 Bribe game - aligned version

In this section, you have been matched with another person. Like you, this person is a participant in the current study. In this section, you have a choice between two options, Option A and Option B. The two options are as follows:

Option A
If you choose Option A, you will help children with cancer. Your choice of Option A will trigger a donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.*
Option B
If you choose option B, you will send the other person 2 EUR. However, in this case, the donation otherwise triggered by option A will not be given to *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.*

Additional consequence

The other person also has the choice between two options, namely between option C and option D. If the other person chooses option C, an additional 20 EUR will be donated to the *Förderkreis* If the person chooses option D, the person will receive 2 EUR for themselves as an additional payout.

The person has chosen to make a choice between the two options, C and D, dependent on **your decision** between A and B. This means the other person has **obligatorily** specified the following: The person will choose option D and thus the additional payout for himself. The choice of option D can only be prevented if you choose option A and thus the donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* . In this and only in this case, the other person will choose option C and thus trigger the donation. Thus, the consequences are as follows:

- You choose option A $\Rightarrow$ The other person chooses option C.
- You choose option B $\Rightarrow$ The other person chooses option D.

In other words, if you choose option A and thus trigger a donation of 15 EUR, the other person will choose option C and thus the 20 EUR donation. If you choose to send the 2 EUR to the other person by choosing option B, the other person will stick to choosing option D and thereby receive 2 EUR.

Decision screen

You can now choose between the two options. Click here if you would like to read the instructions again. As a reminder, here are the two options:

- **Option A:** I help children with cancer by donating 15 EUR to the *Förderkreis*.
- **Option B:** I send 2 EUR to the other person.

Your choice: []

J.2.10 Bribe game - Randomization version

[Instructions before the decision screen are identical to the tradeoff/aligned version]

Decision screen

You can now choose between the two options. Click here if you would like to read the instructions again. As a reminder, here are the two options:

- **Option A:** I help children with cancer by donating 15 EUR to the *Förderkreis*.
- **Option B:** I send 2 EUR to the other person.

Selection of Options

You choose between the two options by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Option A will be implemented with a probability of 0% and Option B with a probability of 100%.
- If you enter 1, Option A will be implemented with a probability of 10% and Option B with a probability of 90%.
- If you enter 2, Option A will be implemented with a probability of 20% and Option B with a probability of 80%.
- And so on, up to the number 10, at which point Option A will be implemented with a probability of 100% and Option B with a probability of 0%.

Decision

Choose a number from 0 to 10: []

Consequence of Your Decision

Option A will be implemented with a probability of *–Enter a number–* %.

Option B will be implemented with a probability of *–Enter a number–* %.

J.2.11 Group decision game - Tradeoff version

Information

Your group

Together with 5 other people, you will form a group for your decisions. These people also participate in this study, so together, you will form a group of 6 individuals.
You can make all decisions autonomously and independently of the other members of the group. However, as described below, the consequences of your decision will depend on the decisions of the other group members.

The donation

Your group has been entrusted with a donation of 15 EUR, which is intended for the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V. and is meant to be donated to the charity after the study.

Your options

You can choose between two options: Option A and Option B. Depending on which option you and the other group members choose, different consequences will be realized. In particular, both your decision and the decisions of the other group members affect what happens to the donation of 15 EUR to the Föderkreis.

The donation to the Föderkreis will be destroyed (that is, not executed) if at least one member of your group chooses option B. That is, if either you or at least one other person with whom you form a group chooses option B, the donation will be destroyed. Only if no member of your group chooses option B, i.e., **all** group members choose option A, will the donation not be destroyed.
To summarize, the consequences of choosing Option A and Option B are as follows:

Option A
You choose to help children with cancer through the donation. However, the donation will only realize if all other members of your group choose Option A.

Option B
You decide against helping children with cancer through the donation. Since the donation will only be made if all group members choose option A, this will destroy the donation in any case.

Furthermore, by choosing option B, you generate an additional payoff of 2 EUR for another participant in this study who is not part of your group. Any other person in your group who chooses option B will receive a sum of money of 2 EUR for themselves.

Decisions of the other group members

Hence, the consequences of your choice depend not only on you but also on the decisions of the other 5 members of your group: only if none of the members of your group chooses option B, will the donation be made. On the other hand, the payouts of 2 EUR to other people will be made regardless of the decisions of other group members.

You and the other 5 Members of your group decide one after the other. You are the **last** person to decide, so the other group members have already decided at this point. When it's your turn, you will learn if anyone among the people who decided before you has already chosen option B. In this case, the donation will be destroyed.

Summary

If you choose option A, you will not generate any additional payout. If you choose option B, you generate an additional payout of 2 EUR for another participant. Whether the donation of 15 EUR to the *Förderkreis* is destroyed depends on whether at least one member of your group has chosen option B. So the consequences of your choice depend not only on you but also on the choices of the other group member.

[Comprehension questions]

Decision screen

Situation: The other people in your group have already made their decisions. At least one person has chosen Option B.
You can now choose between the two options. As a reminder, here are the two options:

- **Option A:** I choose the donation in order to help children with cancer.
- **Option B:** I choose against the donation, in order to send 2 EUR to the other person.

Your choice: [ ]

J.2.12 Group decision game - aligned version

Information

Your group

Together with 5 other people, you will form a group for your decisions. These people also participate in this study, so together, you will form a group of 6 individuals.
You can make all decisions autonomously and independently of the other members of the group. However, as described below, the consequences of your decision will depend on the decisions of the other group members.

The donation

Your group has been entrusted with a payment of 2 EUR, which is intended for another person, who is not part of your group.

Your options

You can choose between two options: Option A and Option B. Depending on which option you and the other group members choose, different consequences will be realized. In particular, both your decision and the decisions of the other group members affect what happens to the payment of 2 EUR to the other person.

The payment will be destroyed (that is, not executed) if at least one member of your group chooses option A. That is, if either you or at least one other person with whom you form a group chooses option A, the donation will be destroyed. Only if no member of your group chooses option A, i.e., **all** group members choose option B, will the donation not be destroyed.
To summarize, the consequences of choosing Option A and Option B are as follows:

Option A
You choose to help children with cancer through the donation. Furthermore, by choosing option A, you generate an additional donation of 15 EUR to the *Förderkreis*. Since the payment to the other person will only be made if all group members choose option B, this will destroy the payment in any case.

Option B
You decide against helping children with cancer through the donation and for the additional payment to another person, who is not part of your group. However, this payment only be made if all other members of your group choose option B.

Decisions of the other group members

Hence, the consequences of your choice depend not only on you but also on the decisions of the other 5 members of your group: only if none of the members of your group chooses option A, will the payment be made. On the other hand, the donation of 15 EUR to the charity will be made regardless of the decisions of other group members.

You and the other 5 Members of your group decide one after the other. You are the **last** person to decide, so the other group members have already decided at this point. When it's your turn, you will learn if anyone among the people who decided before you has already chosen option A. In this case, the payment will be destroyed.

Summary

If you choose option A, you generate a donation of 15 EUR to the *Förderkreis*. If you choose option B, you will not generate any donation. Whether the payment of 2 EUR to another person is destroyed depends on whether at least one member of your group has chosen option A. So the consequences of your choice depend not only on you but also on the choices of the other group member.

[Comprehension questions]

Decision screen

Situation: The other people in your group have already made their decisions. At least one person has chosen Option A.
You can now choose between the two options. As a reminder, here are the two options:

- **Option A:** I choose the donation in order to help children with cancer.
- **Option B:** I choose against the donation, in order to send 2 EUR to the other person.

Your choice: [ ]

J.2.13 Group decision game - Randomization version

[Instructions before the decision screen are identical to the tradeoff/aligned version]

Decision screen

Situation: The other people in your group have already made their decisions. At least one person has chosen Option B.
You can now choose between the two options. As a reminder, here are the two options:

- **Option A:** I choose the donation in order to help children with cancer.
- **Option B:** I choose against the donation, in order to send 2 EUR to the other person.

You choose between the two options by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Option A will be implemented with a probability of 0% and Option B with a probability of 100%.
- If you enter 1, Option A will be implemented with a probability of 10% and Option B with a probability of 90%.
- If you enter 2, Option A will be implemented with a probability of 20% and Option B with a probability of 80%.
- And so on, up to the number 10, at which point Option A will be implemented with a probability of 100% and Option B with a probability of 0%.

Decision

Choose a number from 0 to 10: []

Consequence of Your Decision

Option A will be implemented with a probability of *–Enter a number–* %.

Option B will be implemented with a probability of *–Enter a number–* %.

J.2.14 Statement choice - Tradeoff version

Context

Carbon dioxide (CO_2) is a gas that is notably produced as a by-product of human-controlled combustion processes of fossil fuels such as crude oil, natural gas, or coal. One example of this is the generation of electricity in coal-fired power plants or driving a car. As a result, CO_2 enters the atmosphere, which leads to global warming (greenhouse effect). The increasing CO_2 concentration in the atmosphere, therefore, contributes significantly to man-made climate change and, thus, to the advancing destruction of the environment.

In the context of this study, it was planned to buy CO_2 certificates, which ensure that **one ton of** CO_2 is removed from the atmosphere. More precisely, this entails an amount of 25 EUR, which will be used for reforestation. The reforestation afforded by this amount of money results in the removal of one ton of CO_2 from the atmosphere.

Consequences

At the end of the study, a computer will randomly decide what will happen to the planned CO_2 certificates. The following two possibilities exist:

- With a probability of 50%, the certificates will be purchased as planned, thus helping the environment by removing one ton of CO_2 from the atmosphere.
- With a probability of 50%, the certificates will be destroyed, i.e., not acquired. This contributes to the destruction of the environment by leaving one ton of CO_2 in the atmosphere.

This selection is made by the computer in a completely random fashion. Hence it is not influenced by any actions or decisions of yours at any time.

In this context, you can make a statement of intent about how you feel about the destruction of the environment. Namely, you can make one of the following two statements:

- I support the preservation and protection of the environment.
- I support the destruction of the environment.

If you make the statement “I support the preservation and protection of the environment,” in the case that the certificates are destroyed, a part of the amount will be given to another study participant. Namely, the other participant will receive 2 EUR.

However, if you declare “I support the destruction of the environment” in the case that the certificates are destroyed, a part of the amount will be donated to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* instead. Namely, in this case, 15 EUR will be donated to the *Förderkreis*.

Thus, by selecting a statement, you do not influence whether the CO_2 allowances are destroyed or not (since this happens randomly). Instead, you can influence whether, in the event that the certificates are destroyed, of the released funds 15 EUR will be donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V. or another study participant receives 2 EUR.

Please note that selecting a statement is not an opinion poll. Specifically, regardless of which statement you choose, the results will not be used to influence opinions about the environment.

Decision screen

Reminder: Regardless of your decision, the CO_2 certificates will be destroyed with a probability of 50%. If you state that you support the destruction of the environment (Statement 2), then in the case of destruction of the certificates, an amount of 15 EUR from the released funds will be donated to the support association.
Click here if you would like to review the full instructions again. As a reminder, here are the two statements:

- **Statement 1:** I support the preservation and protection of the environment.
- **Statement 2:** I support the destruction of the environment.

Choose the statement: []

J.2.15 Statement choice - aligned version

Context

Carbon dioxide (CO_2) is a gas that is notably produced as a by-product of human-controlled combustion processes of fossil fuels such as crude oil, natural gas, or coal. One example of this is the generation of electricity in coal-fired power plants or driving a car. As a result, CO_2 enters the atmosphere, which leads to global warming (greenhouse effect). The increasing CO_2 concentration in the atmosphere, therefore, contributes significantly to man-made climate change and, thus, to the advancing destruction of the environment.

In the context of this study, it was planned to buy CO_2 certificates, which ensure that **one ton of CO_2** is removed from the atmosphere. More precisely, this entails an amount of 25 EUR, which will be used for reforestation. The reforestation afforded by this amount of money results in the removal of one ton of CO_2 from the atmosphere.

Consequences

At the end of the study, a computer will randomly decide what will happen to the planned CO_2 certificates. The following two possibilities exist:

- With a probability of 50%, the certificates will be purchased as planned, thus helping the environment by removing one ton of CO_2 from the atmosphere.
- With a probability of 50%, the certificates will be destroyed, i.e., not acquired. This contributes to the destruction of the environment by leaving one ton of CO_2 in the atmosphere.

This selection is made by the computer in a completely random fashion. Hence it is not influenced by any actions or decisions of yours at any time.

In this context, you can make a statement of intent about how you feel about the destruction of the environment. Namely, you can make one of the following two statements:

- I support the preservation and protection of the environment.
- I support the destruction of the environment.

If you make the statement "I support the destruction of the environment," in the case that the certificates are destroyed, a part of the amount will be given to another study participant. Namely, the other participant will receive 2 EUR.

However, if you declare "I support the preservation and protection of the environment" in the case that the certificates are destroyed, a part of the amount will be donated to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* instead. Namely, in this case, 15 EUR will be donated to the *Förderkreis*.

Thus, by selecting a statement, you do not influence whether the CO_2 allowances are destroyed or not (since this happens randomly). Instead, you can influence whether, in the event that the certificates are destroyed, of the released funds 15 EUR will be donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V. or another study participant receives 2 EUR.

Please note that selecting a statement is not an opinion poll. Specifically, regardless of which statement you choose, the results will not be used to influence opinions about the environment.

Decision screen

Reminder: Regardless of your decision, the CO_2 certificates will be destroyed with a probability of 50%. If you state that you support the preservation and protection of the environment (Statement 1), then in the case of destruction of the certificates, an amount of 15 EUR from the released funds will be donated to the support association.
Click here if you would like to review the full instructions again. As a reminder, here are the two statements:

- **Statement 1:** I support the preservation and protection of the environment.
- **Statement 2:** I support the destruction of the environment.

Choose the statement: []

J.2.16 Statement choice - Randomization version

[Instructions before the decision screen are identical to the tradeoff/aligned version]

Decision screen

Reminder: Regardless of your decision, the CO_2 certificates will be destroyed with a probability of 50%. If you state that you support the destruction of the environment (Statement 2), then in the case of destruction of the certificates, an amount of 15 EUR from the released funds will be donated to the support association.
Click here if you would like to review the full instructions again. As a reminder, here are the two statements:

- **Statement 1:** I support the preservation and protection of the environment.
- **Statement 2:** I support the destruction of the environment.

Selection of Statements

You choose between the two statements by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, you make Statement 1 with a probability of 0% and Statement 2 with a probability of 100%.
- If you enter 1, you make Statement 1 with a probability of 10% and Statement 2 with a probability of 90%.
- If you enter 2, you make Statement 1 with a probability of 20% and Statement 2 with a probability of 80%.
- And so on, up to the number 10, at which point you make Statement 1 with a probability of 100% and Statement 2 with a probability of 0%.

Decision

Choose a number from 0 to 10:

Consequence of Your Decision

You make Statement 1 with a probability of *–Enter a number–* %.

You make Statement 2 with a probability of *–Enter a number–* %.

J.2.17 Rule following game - Tradeoff version

On the next screen, you will control a character that runs across the screen in an animation.
As soon as the animation starts, you can start running by pressing the "Start" button in the middle of the screen. Your character will move towards a series of traffic lights and will wait at each one. To move your character again, press the "Walk" button in the middle of your screen.
The rule is that you should wait at each traffic light until it turns green.
Your payout depends on how much time you need to run across the screen with your character. You start with an initial payout of 8 EUR, which decreases by 0.08 EUR every second.
The next screen takes you to the animation. If the animation is too large for your browser window, please change the zoom factor so that the animation is displayed in its entirety.

J.2.18 Rule following game - aligned version

On the next screen, you will control a character that runs across the screen in an animation.
As soon as the animation starts, you can start running by pressing the "Start" button in the middle of the screen. Your character will move towards a series of traffic lights and will wait at each one. To move your character again, press the "Walk" button in the middle of your screen.
The rule is that you should wait at each traffic light until it turns green.
Your payout depends on how much time it takes you to walk your character across the screen. You start with an initial payout of 0 EUR. Your payout increases by 0.08 EUR every second, up to a maximum of 8 EUR. The longer you wait at the traffic lights, the higher your payout will be. You will reach the maximum payout if you wait at each traffic light until it turns green.
The next screen takes you to the animation. If the animation is too large for your browser window, please change the zoom factor so that the animation is displayed in its entirety.

J.3 Self-versus-others block

J.3.1 Dictator game giving - Main version

In this section, you will receive 20 EUR as an additional payment. If you take no further action, you will receive the full amount. You can choose to donate a certain amount of the 20 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* You can choose any amount between 0 EUR and 20 EUR. This amount will then not be paid out to you but donated to the *Förderkreis*.

Thus, you will receive an additional payment 20 EUR minus the amount you donate.

Decision screen

Your choice: []

J.3.2 Dictator game giving - Randomization version

In this section, you receive an additional payment of 20 EUR. You can decide to donate a certain amount of the 20 EUR to the *F́orderkreis* fÃŒr krebskranke Kinder und Jugendliche Bonn e.V.

- **Option A:** You choose not to donate any money from the 20 EUR. You therefore receive 20 EUR as an additional payment, and no money is donated.
- **Option B:** You choose to donate 10 EUR of the 20 EUR to the association. You therefore receive 10 EUR as an additional payment, and 10 EUR will be donated.

Decision screen

You can now choose between the two options. As a reminder, here are the two options:

- **Option A:** You choose not to donate any money from the 20 EUR. You therefore receive 20 EUR as an additional payment, and no money is donated.
- **Option B:** You choose to donate 10 EUR of the 20 EUR to the association. You therefore receive 10 EUR as an additional payment, and 10 EUR will be donated.

Selection of Options

You choose between the two options by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Option A will be implemented with a probability of 0% and Option B with a probability of 100%. In this case, Option B will definitely be implemented.
- If you enter 1, Option A will be implemented with a probability of 10% and Option B with a probability of 90%.
- And so on, up to the number 10, at which point Option A will be implemented with a probability of 100% and Option B with a probability of 0%. In this case, Option A will definitely be implemented.

Decision

Choose a number from 0 to 10:

Consequence of Your Decision

Option A will be implemented with a probability of –*Enter a number*– %.

Option B will be implemented with a probability of –*Enter a number*– %.

J.3.3 Dictator game taking - Main version

This section is about a donation of 20 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* If you take no further action, the full amount will be donated. You can choose to take a certain amount of the donation for yourself as an additional payment. You can choose any amount between 0 EUR and 20 EUR. This amount will then not be donated to the *Förderkreis*, but paid to you.
Thus, 20 EUR minus the amount you take for yourself will be donated.

Decision screen

Your choice: []

J.3.4 Trust game sender - Main version

Information

In this section, you and another person can send money to each other. Both of you have a private account. Initially, you and the other person have an endowment of 5 EUR in the account. Like you, the other person is a participant in today's study.

Your decision

From your 5 EUR endowment, you can decide to send an amount to the other person. You can send any whole euro amount, that is 1,2,3,4 or 5 EUR to the other person. In this case, each euro you send to the other person will arrive tripled to the other person. So, for example, if you send 1 EUR, the other person will receive 3 EUR, if you send 2 EUR, 6 EUR and so on.

The decision of the other person

Thus, after you make your decision, the other person has 5 EUR + triple the amount you sent. Now the other person decides on an amount of money to send back to you. The person can choose any amount that the account allows for. After that, the process ends and you will receive your account balances as an additional payment.

Example

You have decided to send 2 EUR to the other person. Therefore, your account balance is 5 EUR - 2 EUR = 3 EUR. The person now has 5 EUR + 3 X 2 EUR = 11 EUR on the account and can accordingly send you back any amount between 0 EUR and 11 EUR. Suppose the person sends you back 4 EUR.

This will give you a total of 4 EUR + 3 EUR = 7 EUR as an additional payment. The other person will receive 11 EUR - 4 EUR = 7 EUR as a payment.

J.3.5 Trust game sender - Randomization version

In this section, you can send an amount of money to another person. Afterwards, the other person can decide whether to send an amount of money back to you.
You both have a private account. At the beginning, both you and the other person have an endowment of 5 EUR in your accounts. The other person, like you, is a participant in today's study.

Your Decision

From your endowment of 5 EUR, you can decide whether to send an amount to the other person. You have two options, Option A and Option B:

- **Option A:** You do not send any money to the other person. You therefore keep your endowment of 5 EUR as an additional payment.
- **Option B:** You send 2 EUR to the other person. This amount is tripled, so the other person receives 6 EUR.

The Other Person's Decision

After you have made your decision, the other person has 5 EUR plus the tripled amount you may have sent.

- If you chose Option A: the other person has $5 + 0 = 5$ EUR.
- If you chose Option B: the other person has $5 + 6 = 11$ EUR.

Now the other person decides whether to send money back to you. They also have two options:

- **Option 1:** Send nothing back to you.
- **Option 2:** Send 4 EUR back to you.

Afterwards, the process ends, and you both receive your respective account balances as additional payments.

Possible Outcomes

If you choose Option A: Your intermediate balance is 5 EUR.

- If the other person chooses Option 1 (send nothing), you receive $5 + 0 = 5$ EUR. The other person receives $5 - 0 = 5$ EUR.

- If the other person chooses Option 2 (send 4 EUR), you receive $5 + 4 = 9$ EUR. The other person receives $5 - 4 = 1$ EUR.

If you choose Option B: Your intermediate balance is 3 EUR.

- If the other person chooses Option 1 (send nothing), you receive $3 + 0 = 3$ EUR. The other person receives $5 + 6 = 11$ EUR.
- If the other person chooses Option 2 (send 4 EUR), you receive $3 + 4 = 7$ EUR. The other person receives $5 + 6 - 4 = 7$ EUR.

Decision screen

You can now choose between the two options. As a reminder, here are the two options:

- **Option A:** You choose not to send any money from your endowment of 5 EUR to the other person.
- **Option B:** You choose to send 2 EUR from your endowment of 5 EUR to the other person.

Selection of Options

You choose between the two options by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Option A will be implemented with a probability of 0% and Option B with a probability of 100%. In this case, Option B will definitely be implemented.
- If you enter 1, Option A will be implemented with a probability of 10% and Option B with a probability of 90%.
- And so on, up to the number 10, at which point Option A will be implemented with a probability of 100% and Option B with a probability of 0%. In this case, Option A will definitely be implemented.

Decision

Now choose a number from 0 to 10:

Consequence of Your Decision

Option A will be implemented with a probability of –*Enter a number*– %.

Option B will be implemented with a probability of –*Enter a number*– %.

J.3.6 Trust game receiver - Main version

Information

In this section, you now take the other person's role in the previous section. That is, now the other person decides how much money to send you, and you decide how much to send back. Note that you are sending money to a different participant for this section than before. They are not the same person, and your decision in the previous section has no bearing on your decision in this section.

Both of you again have a private account with an endowment of 5 EUR. Now the other person can decide how much of the 5 EUR they want to send you.

Your decision

You can indicate what you would send back for different possible amounts the other person can send. Afterward, it will be checked what amount the other person actually sent, and your decision for exactly this case will be implemented.

J.3.7 Trust game receiver - Randomization version

In this section, you now take on the role that the other person held in the previous section. This means that the other person first decides how much money to send to you, and you then decide how much to send back.
Note that in this section, you are paired with a different participant than before. It is not the same person, and your decision in the previous section has no influence on this one.
You both again have a private account, each with an endowment of 5 EUR. The other person now decides how much of their 5 EUR to send to you.

Your Decision

You now have the choice, for each of the two possible amounts the other person may send, to state how much you would return. In each case, you can choose between two options:

- **Option 1:** You send nothing back.
- **Option 2:** You send 4 EUR back.

You will make two decisions:

- The first decision applies to the case where the other person sends you nothing.
- The second decision applies to the case where the other person sends you 2 EUR (which will be tripled).

After you have made both decisions, the amount that the other person actually sent will be revealed, and the corresponding decision you made will be implemented.

Decision screen 1

Decide now for the case in which the other person has sent you no money. You therefore have 5 EUR in your account. How much would you like to send back to the other person?

- **Option 1:** You choose not to send any money back.
- **Option 2:** You choose to send 4 EUR back.

Selection of Options

You choose between the two options by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Option 1 will be implemented with a probability of 0% and Option 2 with a probability of 100%. In this case, Option 2 will definitely be implemented.
- If you enter 1, Option 1 will be implemented with a probability of 10% and Option 2 with a probability of 90%.
- And so on, up to the number 10, at which point Option 1 will be implemented with a probability of 100% and Option 2 with a probability of 0%. In this case, Option 1 will definitely be implemented.

Decision

Now choose a number from 0 to 10:

Consequence of Your Decision

Option 1 will be implemented with a probability of –*Enter a number*– %.

Option 2 will be implemented with a probability of –*Enter a number*– %.

Decision screen 2

Decide now for the case in which the other person has sent you 2 EUR. This amount is tripled, so you have $5 + 6 = 11$ EUR in your account. How much would you like to send back to the other person?

- **Option 1:** You choose not to send any money back.
- **Option 2:** You choose to send 4 EUR back.

Selection of Options

You choose between the two options by entering a number from 0 to 10 in the text field below. Your input has the following consequences:

- If you enter 0, Option 1 will be implemented with a probability of 0% and Option 2 with a probability of 100%. In this case, Option 2 will definitely be implemented.
- If you enter 1, Option 1 will be implemented with a probability of 10% and Option 2 with a probability of 90%.
- And so on, up to the number 10, at which point Option 1 will be implemented with a probability of 100% and Option 2 with a probability of 0%. In this case, Option 1 will definitely be implemented.

Decision

Now choose a number from 0 to 10:

Consequence of Your Decision

Option 1 will be implemented with a probability of –*Enter a number*– %.

Option 2 will be implemented with a probability of –*Enter a number*– %.

J.3.8 Public goods game - Main version

Your group

In this section, you will form a group with 2 other people that also participate in today's study. Your group members have been randomly assigned to you. At no time will you or the other members of your group learn who the members of your group are. All group members will face the same decisions and receive the same information.

Your endowment

For this section, you and each other group member will receive an amount of 5 EUR which is called your endowment. You decide how to use your endowment. You can contribute all or part of it to a project. Any amount you do not contribute to the project will automatically be put into what is called a private account. For example, if you contribute 3 EUR to the project, 5 EUR- 3 EUR = 2 EUR will be put into your private account.

You will receive income from the project and your private account as described below.

Income from the private account

For every Euro that is put into your private account, you will receive exactly 1 EUR as income. For example, if you contribute 0 EUR of your endowment to the project and thus 5 EUR is put into your private account, you will receive 5 EUR as income from your private account. If 3 EUR is put into your private account (because you contributed 2 EUR to the project), you will receive 3 EUR as private income. Nobody but you will receive income from your private account.

Income from the project

For every EUR you or another group member contributes to the project, you and all other group members receive 0.50 EUR. So, each group member's income from the project is determined as follows:

Income from the project = 0.5 * total sum of contributions paid in by all group members.

Example: If the total sum of contributions to the project from all participants in your group is 9 EUR (for example, if you and the two other members each contribute 3 EUR), you and all other participants in your group will receive 9 EUR * 0.5 = 4.5 EUR from the project. If the total of your contributions to the project is 12 EUR, you and all other participants will receive 12 EUR * 0.5 = 6 EUR from the project.

Total income

Your total income from part 1 is simply the sum of your income from the private account and your income from the project.

Simulation

You can try out in the simulation below how your total income changes based on your contribution and the contributions of your group members to the project. To do this, enter whole numbers in the respective fields.

[Calculator]

[Comprehension questions]

J.3.9 Public goods game - Randomization version

In this section, you form a group of three people together with two other participants in today's study. Your group members have been randomly assigned. Neither you nor the other members of your group will ever know who the group members are. All group members make the same decisions and receive the same information.

Your Endowment

For this section, you and each group member receive an amount of 5 EUR, referred to as your endowment. You decide how to use your endowment and have two options: You can either contribute 0 EUR to a project (Option A) or contribute 3 EUR to the project (Option B).
The amount that you do not contribute to the project is automatically placed into a so-called private account. Thus, if you contribute 3 EUR to the project, then $5 - 3 = 2$ EUR is placed in your private account. If you contribute 0 EUR to the project, then $5 - 0 = 5$ EUR is placed in your private account.

You receive income from both the project and your private account, as described below.

Income from the Private Account

There are two possibilities:

- **Option A:** If you contribute 0 EUR of your endowment to the project, then 5 EUR is placed into your private account, and you receive 5 EUR as income from your private account.
- **Option B:** If you contribute 3 EUR to the project, then 2 EUR is placed into your private account, and you receive 2 EUR as income from your private account.

No one other than you receives income from your private account.

Income from the Project

For every 1 EUR contributed to the project by you or another group member, you and all other group members receive 0.50 EUR. Thus, the income of each group member from the project is determined as follows:

$$\text{Income from Project} = 0.5 \times \text{Total Contributions of All Group Members}$$

Examples

- If the sum of contributions of all group members is 9 EUR (i.e., you and the other two members all choose Option B), then you and all other group members receive $9 \times 0.5 = 4.5$ EUR from the project.
- If the sum of contributions of all group members is 3 EUR (i.e., two members choose Option A and one member chooses Option B), then you and all other group members receive $3 \times 0.5 = 1.5$ EUR from the project.
- If the sum of contributions of all group members is 0 EUR (i.e., all members choose Option A), then you and all other group members receive $0 \times 0.5 = 0$ EUR from the project.

Total Income

Your total income is simply the sum of your income from the private account and your income from the project.

Your Decision

You can choose between two options, Option A and Option B:

- **Option A:** You contribute nothing to the group project. You therefore place your full endowment of 5 EUR into your private account.

- **Option B:** You contribute 3 EUR to the group project. You therefore place 2 EUR into your private account.

[Calculator]

Decision screen

You can now choose between the two options. As a reminder, here are the two options:

- **Option A:** You choose not to contribute any money from your endowment of 5 EUR to the project.
- **Option B:** You choose to contribute 3 EUR from your endowment of 5 EUR to the project.

Selection of Options

You choose between the two options by entering a number from 0 to 10 in the lower text field. Your input has the following consequences:

- If you enter 0, Option A will be implemented with a probability of 0% and Option B with a probability of 100%. In this case, Option B will definitely be implemented.
- If you enter 1, Option A will be implemented with a probability of 10% and Option B with a probability of 90%.
- And so on, up to the number 10, at which point Option A will be implemented with a probability of 100% and Option B with a probability of 0%. In this case, Option A will definitely be implemented.

Decision

Now choose a number from 0 to 10:

Consequence of Your Decision

Option A will be implemented with a probability of –*Enter a number*– %.

Option B will be implemented with a probability of –*Enter a number*– %.

J.3.10 Moral luck game

Information

The following is about the decisions of another person who, like you, is a participant in this study. However, this person faced a different decision than you will be presented. At an earlier time, the person had a choice between the following two lotteries S and G:

Lottery S – S for Donation [Spende in German].

With a probability of 70%, a donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* is triggered. With a probability of 30%, the person receives 10 EUR as an additional payment and no donation is triggered.

Lottery G – G for Money [Geld in German].

With a probability of 30%, a donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* is triggered. With a probability of 70%, the person receives 10 EUR as an additional payment and no donation is triggered.

Determination of the consequences

At the end of the trial, the computer will randomly draw a number for the person from numbers 1 to 10. Depending on which lottery the person chose, the consequences of the number drawn will be different.

If the person chose lottery S, the donation will be triggered with a probability of 70%, so if either a 1, 2, 3, 4, 5, 6, or 7 is drawn. If the number drawn is an 8, 9, or 10, the person will receive 10 EUR as an additional payment.

If the person chose lottery G, the donation is triggered with a probability of 30%, so if either a 1, 2, or 3 is drawn. If the number drawn is a 4, 5, 6, 7, 8, 9, or 10, the person will receive 10 EUR as an additional payment.

Thus, the two lotteries differ in their probabilities of triggering the donation and the amount of additional payment to the other person. For Lottery S, a donation is much more likely than for Lottery G. Conversely, the probability of receiving an additional payment is higher for Lottery G than for Lottery S.

The consequences of the lotteries have no influence on your own payments.

Your decisions

In this section you will make decisions for different scenarios. For your decisions, you will be given 5 EUR for each scenario. You have the choice to divide the 5 EUR between you and the person who decided between the lotteries. You will receive every Euro you allocate to yourself as an additional payment. For every Euro you allocate to the other person, the person will subsequently receive 3 EUR as an additional payment.

For example, you can divide the 5 EUR so that you receive 4 EUR. Accordingly, the other person will then receive 1 EUR * 3 = 3 EUR as an additional payment. If you allocate yourself the entire 5 EUR, the other person will receive 0 EUR. If you allocate the entire amount to the other person you will receive 0 EUR and the other person will receive 5 EUR * 3 = 15 EUR and so on.

You will be asked how you want to distribute the amount of money from 5 EUR for each of four different scenarios.

Scenarios

The scenarios differ in which lotteries the person chose in each case, as well as which consequence happened to materialize. The following four scenarios exist:

Scenario 1:
The other person chose Lottery S. The donation was triggered and the person does not receive the additional payment.

Scenario 2:
The other person chose Lottery S. The donation was not triggered and the person receives the additional payment.

Scenario 3:
The other person chose Lottery G. The donation was triggered and the person does not receive the additional payment.

Scenario 4:
The other person chose Lottery G. The donation was not triggered and the person receives the additional payment.

For each of the four scenarios, you can decide individually how to distribute the 5 EUR based on the person's decision and the realized consequences. At no point does the other person receive any information about you.

At the end of the trial, you will be informed which scenario occurred by the matching of a person's decision to the consequences the computer has drawn for them.

Afterwards, you and the person then receive the amounts of money you selected for the scenario that occurred.

On the next screen you will see an example.

Example

You will see an example decision for Scenario 1. You can make the distribution of 5 EUR between you and the person using a scale and a slider. The associated slider for the scale appears when you click on the scale.

[Example]

If scenario 1 occurs, you would receive [XX] EUR by your decision and the other person would receive [XX] EUR. The other person will receive a total of [XX] EUR, because in this scenario the realized consequence is that the other person will not receive an additional amount of money through the lottery, but will have triggered a donation.