Disentangling Model-Based and Model-Free Moral Learning

Zahra Tahmasebi[1]*, Maximilian Maier[2]*, Vanessa Cheung[2]*, Fiery Cushman[3], & Falk Lieder[1]
[1] Department of Psychology, University of California, Los Angeles, CA, 90095, USA
[2] Department of Experimental Psychology, University College London, 26 Bedford Way, WC1H 0AP, London, UK
[3] Department of Psychology, Harvard University, Cambridge, MA, 02139, USA
* These authors contributed equally.

Abstract

To resolve moral dilemmas, people often rely on decision strategies such as cost-benefit reasoning (CBR) or following moral rules. Previous studies show that people learn to increasingly rely on whichever strategy led to better outcomes in the past. Do they learn this by constructing a mental model of what outcomes would result from using either strategy (i.e., model-based learning) or by assigning value directly to each strategy (i.e., model-free learning)? To answer this question, we adapted the two-step task to a trolley-type dilemma between following moral rules (e.g., obeying authority) versus CBR (e.g., saving a larger group). In each of the 125 trials, participants' choices led to either a common or a rare transition, which probabilistically led to good versus bad outcomes. Computational modeling and pre-registered analysis of behavioral data provided converging evidence that participants apply both model-based and model-free learning.

Keywords: moral decision-making; two-step task; moral dilemmas; moral learning; metacognitive learning

Introduction

In *Les Misérables*, Jean Valjean is faced with a dilemma. He must decide whether to own up to a past crime when another man is falsely accused of it, or to stay silent and do more good in his new life as a mayor, where he can help workers in poverty. This is a dilemma between following the moral rule of telling the truth or violating the rule for what he perceives to be the greater good. Dilemmas like these are common both in high-stakes societal decisions (e.g., whether to implement lockdowns during a pandemic) and in everyday decisions with lower stakes (e.g., whether to compliment someone's outfit when you do not like it).

Jean Valjean's dilemma highlights that people can use different strategies to make decisions in moral dilemmas, such as following moral rules (e.g., do not kill) or relying on cost-benefit reasoning (CBR; e.g., weigh the cost of lying against its benefits). These decision strategies are reminiscent of the ethical theories of deontology and consequentialism[1], respectively, but they are not equivalent. Ethical theories specify normative criteria for what is morally right or wrong that are often impractical to apply in everyday life. In contrast, a decision strategy is a practical step-by-step process that selects actions that may or may not be optimal according to the person's ethical theory. For example, the ethical theory of consequentialism states that one should choose the action with the best consequences. However, the decision strategy of CBR often falls short of selecting such actions. Moreover, following moral rules, a decision strategy reminiscent of the ethical theory of deontology, can be viewed as a heuristic for selecting the actions with the best consequences (Maier et al., 2024; Williams, 2023), and can lead to better consequences than CBR in environments where people are often mistaken about what the consequences of their actions will be (Bennis et al., 2010; Gigerenzer, 2008).

Beyond decision *strategies*, the reinforcement learning (RL) framework posits two learning *systems*: the model-based system and the model-free system (Daw et al., 2005). The model-based system builds a probabilistic model of how likely different consequences will occur depending on the chosen option. Based on this internal representation, it considers different options, predicts their outcomes, and selects the option that produces the best expected outcomes overall. Conversely, the model-free system assigns value directly to each option based on its history of reward and punishment. The option's value representation is adjusted through prediction error learning by comparing the previous representation of the value with the current outcome.

In most applications of RL, the options are behaviors, but in our application the options are decision strategies. For example, imagine that yesterday a friend asked you about their outfit choice for a party. You chose to tell the truth. Applying this decision strategy led you to tell them their outfit was out of style (behavior). Before they could respond, your friend got interrupted by an unpleasant text message. When they looked back at you, they were upset. Today, your spouse asks you how you like their cooking. Model-free RL from the negative outcome of telling the truth yesterday would make you less likely to use this strategy again. In contrast, the mental model you acquired through model-based RL might tell you that your friend's mood was likely unrelated to your truth-telling because they received an upsetting text.

Depending on the situation, people's decisions seem to follow one system or the other (Gawronski et al., 2017). Previous work with RL perspective on moral decision-making (Cushman, 2013; Crockett, 2013) has linked the model-based system to CBR (e.g., saving more lives; Patil et al., 2020) and the model-free system to following rules (e.g., harm aversion; Lockwood et al., 2020). These studies investigated moral decision-making at the behavioral level and were concerned with the question of which system determines the actions im-

[1]Deontological moral principles often emphasize absolute rules or duties (Alexander & Moore, 2021; Kant & Schneewind, [1785] 2002) and consequentialist theories derive the moral status of an action from its expected consequences. Utilitarianism is a specific form of consequentialism which holds that a moral decision should impartially maximize expected aggregate welfare (Bentham, 1789; Mill, 1879).

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

plied by the strategies of following CBR or rules (i.e., deciding about specific actions). However, it does not answer the separate question of how people choose to follow a certain decision strategy in the first instance (i.e., strategy selection learning; Lieder and Griffiths, 2017) and whether this metacognitive process would be model-based or model-free.

Referring back to the outfit example, deciding whether or not to compliment your friend's outfit is an action-level decision. In contrast, a strategy-level decision would be which general mechanism to employ to make the decision, such as whether to prioritize being honest or being kind.

How do people learn when to rely on which moral decision-making strategy? Maier et al. (2024)'s study on moral learning showed that the consequences of relying on a certain strategy in previous moral decisions increase or decrease the probability of repeating that strategy in the future, depending on whether they were better or worse than expected. This metacognitive learning generalized beyond the experimental paradigm to an incentive-compatible real-life donation decision in a separate experiment. However, it remains an open question whether this learning was model-based (by constructing a model linking strategies to outcomes) or model-free (good/bad outcomes reinforced the strategies directly). This is chiefly because the experimental paradigm was not optimized for disentangling the two types of learning.

To address this gap in the current understanding of moral learning, we sought to disentangle model-free from model-based moral learning from the consequences of past decisions by developing a new experimental paradigm that is optimized for answering that very question: the moral two-step task. In this article, we present this paradigm, use it to conduct an experiment, and analyze the data using computational models of model-free and model-based moral learning.

The Moral Dilemma Two-Step Task

Within the RL framework, the two-step task is the dominant method for dissociating model-based and model-free learning (the task presented in this paper combines features from versions proposed by Daw et al., 2011 and Kool et al., 2016). Each trial consists of a sequence of two decision stages. In the first-stage state, participants choose from two options. Each choice leads to one of two second-stage states. Participants then press a button to reveal the second-stage state, which is determined by a probabilistic reward function that changes during the task. The transition from the first-stage choice to the second-stage state is also probabilistic, but the probabilities are fixed during the experiment. Each first-stage choice has a high probability (e.g., 70%) of transitioning to one of the two sets of options (common transition) and a low probability (e.g., 30%) of transitioning to the other set of options (rare transition).

The design of this task ensures that model-free and model-based learning have opposite effects on the probability that the first decision will be repeated after a rare transition led to a reward. Model-free learning increases the probability that the first decision will be repeated, but model-based learning reduces the probability that it will be repeated. This is because model-free learning, which relies on an action's average consequences in the past, would disregard the transition type and simply repeat the successful action in the first-stage state which ultimately led to a good outcome and vice versa. For example, in the outfit example provided earlier, a model-free learner would not repeat their first-stage decision (being honest) if it led to bad outcome (the friend's mood change), disregarding the fact that the outcome occurred not because of a common transition (the friend reacting to your feedback) but because of a rare transition (an unpleasant text message).

The two-step task had previously been applied to tasks involving moral judgment or behavior. For instance, Patil et al. (2020) used this paradigm to investigate whether model-based learning is associated with utilitarian reasoning, and Lockwood et al. (2020) investigated whether learning of harm aversion is model-free. Both studies adapted the paradigm to the context of moral decision-making by changing the reward state to inducing or avoiding physical harm (loud noise or electric shock) to another person. Apart from this, the decision states were very similar to the non-moral versions of the task: decisions in the first-stage state (e.g., which symbols to select) had no resemblance to moral decisions.

These types of first-stage decisions do not capture relevant features of making a moral decision. Moral decision-making usually involves choosing between two contrasting options in a dilemma (e.g., tell the truth vs. lie but help more people). Therefore, we developed a novel version of the two-step task in the context of moral decision-making. In this task, the first stage confronts people with a moral dilemma between following moral rules versus CBR.

In the moral dilemma two-step task, participants are faced with a moral dilemma in the first-stage state with two conflicting choices: one endorsed by following moral rules and one endorsed by CBR. Each choice leads to a common (70%) versus rare (30%) transition into a second-stage state. The two possible second-stage states lead to good versus bad outcomes probabilistically. We measured participants' first-stage choices, namely the probability of choosing the same option as they did in the previous trial (i.e., "stay probability").

Following the logic of Daw et al. (2011), in our task, if moral learning were model-free, participants would increase their reliance on CBR/rules whenever it led to a good outcome, regardless of the transition type. If moral learning is model-based, then participants would increase their reliance on CBR/rules only if the good outcome resulted from a common transition but decrease it if the good outcome resulted from a rare transition.

Experiment

Method

We preregistered our experiment and data analysis on AsPredicted (#185758). All materials are available on OSF.

Participants This experiment received ethical approval from the Office of the Human Research Protection Program at UCLA under protocol number IRB #23-001436.

We recruited 150 participants from a U.K. and U.S. sample on Prolific on August 8, 2024, based on an a priori power analysis which indicated that 100 participants are required. The median duration for the study was 44 minutes. We excluded participants whose responses were not fully recorded due to a technical issue ($N = 8$). As preregistered, we also excluded participants who did not pass the second attempt at a comprehension check after reading the task instructions ($N = 8$). Of those who passed the comprehension check, we excluded those who failed two or more attention checks during the main task ($N = 4$). Our final sample size was $N = 130$ ($M_{age} = 37.23$, $SD_{age} = 11.96$, $N_{female} = 63$, $N_{male} = 67$).

Design and Materials Participants made 125 repeated decisions. As shown in Figure 1, in the first-stage decision state, we presented a trolley-type moral dilemma. Participants imagined themselves as an employee of a railroad company whose job is to oversee railway junctions where two tracks diverge. A runaway wagon containing dangerous pathogens and explosive materials is approaching the junction. The participant cannot stop the wagon, but must decide whether to direct it to the left or right track by pressing F or J on their keyboard. Unlike classic trolley-type dilemmas, where one must decide between acting to divert the wagon to an alternative track or doing nothing and letting it stay on the default track, we designed the decision to be between switching to a left or a right track to avoid confounding one choice with action and the other with inaction (see Crone and Laham, 2017).

In all trials, the left track leads to a water reservoir that supplies water to a city. If the wagon lands in the reservoir, it would likely explode and leak pathogens into the water, causing a serious disease in about 100 people in the city with a mortality rate of $\sim 2\%$. However, there is a chance that the wagon would remain intact, in which case the water would stay clean. The right track has been closed by the railroad authorities, as indicated by a track closure sign. A worker, who is a colleague of the decision-maker, is conducting maintenance work on the track. If the wagon heads towards him, he would likely be killed. However, there is a chance that he would see the wagon and escape. Participants were also informed that their boss had instructed that railroad company employees have a duty to always obey the track closure signs.

Diverting the wagon to the water reservoir is the option endorsed by moral rules (i.e., the *rule option*), because it follows the moral principle "do not kill", as well as the moral obligation to protect one's colleagues and obey the rules set by an authority figure. Diverting the wagon to the track closure is the option endorsed by CBR (i.e., the *CBR option*) because fewer people would be sacrificed. We had previously piloted these choices to ensure that they are well-balanced, meaning that around 50% of people would prefer each option. The pilot study ($N = 60$) revealed that $M = 45\%$, 95% CI [32%-58%] preferred the rule option, and $M = 55\%$, 95% CI [42%-68%] preferred the CBR option. Further, to check whether participants indeed perceived diverting the wagon towards the track closure to be the option endorsed by CBR, and diverting the wagon towards the water reservoir as the option endorsed by following moral rules, we asked four questions where participants indicated which track would someone who always (1) followed rules, (2) followed their gut feelings, (3) considered the best consequences, and (4) followed reasoning, would choose. As expected, the majority of participants selected the rule option for questions (1) and (2) (70% & 67%),

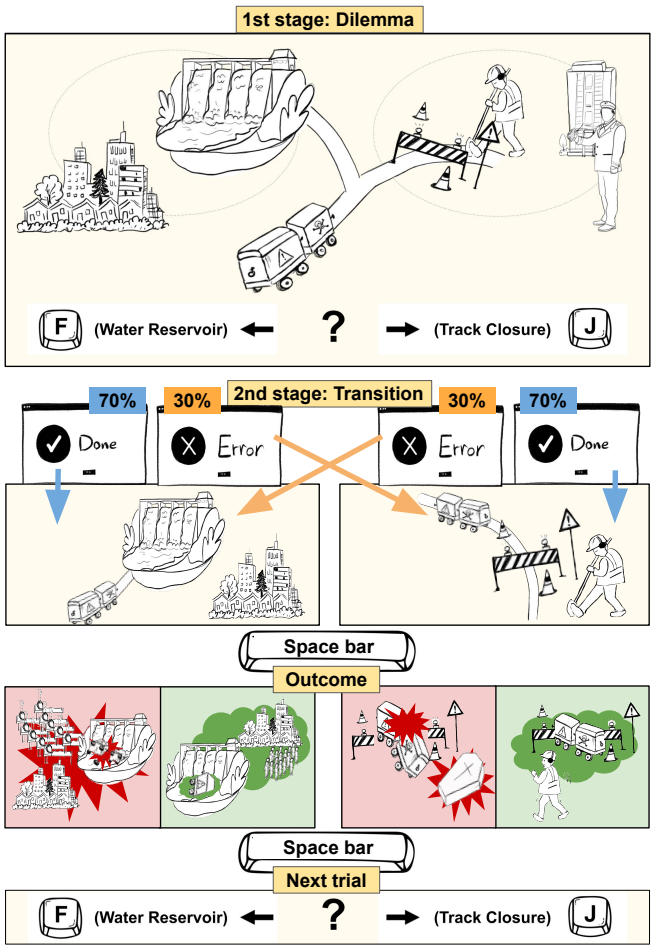

Figure 1: In the moral dilemma two-step task, the first-stage decision state had two conflicting choices: one endorsed by following moral rules (saving your colleague by diverting the wagon to the left track, risking the contamination of a water reservoir causing a disease with 2% mortality rate), and one endorsed by CBR (saving more lives by diverting the wagon to the right track, which is prohibited by authorities and risks killing your colleague). Each of these choices led to a common (70%; wagon moving as directed) versus rare (30%; wagon turning in the opposite direction due to a system error) transition into a second-stage state. The second-stage state then led to good versus bad outcomes probabilistically.

and the CBR option for questions (3) and (4) (74% & 71%).

Each choice in the first-stage moral dilemma led to a common (70%) versus rare (30%) transition into a second-stage state. As part of the cover story, participants were told that every time they choose which track to divert the wagon to, their choice is supplied to a computer system that executes their decision; however, this system is unreliable and sometimes mistakenly diverts the wagon to the opposite (unchosen) track. In the task, there was a 30% chance that when participants chose to divert the wagon to one track, it would actually be diverted to the other (i.e., rare transition to the second-stage state). When this happened, they saw a symbol on the screen indicating a system error and that the wagon went to the opposite track. However, most of the time, the computer worked properly (i.e., common transition), and the participant saw a symbol indicating that their order went through successfully and that the wagon went to the specified track (see Figure 1).

The second-stage state then led to good versus bad outcomes probabilistically. Either a good or a bad outcome could occur regardless of which track the wagon was diverted to. If the wagon was diverted to the water reservoir, it either remained intact, and nobody was harmed (good outcome), or it exploded, and 100 residents got a disease with a 2% mortality rate (bad outcome). If the wagon was diverted to the closed track, either the colleague escaped the track, and nobody was harmed (good outcome), or the wagon ran over the colleague and killed him (bad outcome). The reward probabilities of the second-stage choices (i.e., whether the outcome is good or bad) changed according to a Gaussian random walk ($\mu = 0, sd = 0.025$), initialized at .5 in the first trial and with limits imposed on its range (between 0.25 and 0.75).

Procedure The experiment was programmed on the online platform lab.js. After giving informed consent, participants first read instructions describing the task, including the two choice options, the rare and common transitions, and the positive and negative outcomes that could occur as a result. As part of the instructions, participants completed three rounds of practice trials, with ten trials each. We included three comprehension check questions after the instructions and before the main task. Participants who responded incorrectly to at least one question were required to review the instructions and make a second attempt. Participants who failed any of the comprehension check questions on their second attempt were excluded from analysis.[2]

Participants then took part in the main task, which included 125 trials. We included eight attention checks throughout the task asking about the outcome of their previous decision.

At the end of the task, as exploratory measures, we asked participants a series of questions about how they understood and interpreted the task. We also asked participants if they thought the changes in the study were random or systematic changes, whether the outcomes of previous trials influenced their decision in the next trial, and how frequently they thought the system error (i.e., rare transitions) occurred.

Lastly, participants answered several questionnaires to investigate which factors explain reliance on model-based versus model-free metacognitive moral learning. Specifically, participants indicated how morally right it was to direct the trolley toward either the (1) colleague or (2) the reservoir and completed various scales relevant to moral decision-making: the Empathic Concern (EC) and Emotional Empathy (EE) Scale (Jordan et al., 2016), The Deontology-Consequentialist Deontology Subscale (Mata et al., 2022), the Sacrificial Harm Subscale of the Oxford Utilitarianism Scale (OUS) (Kahane et al., 2018), and the Cognitive Reflection Test (CRT; Frederick, 2005).

Results

Choice Behavior On average, participants more often redirected the wagon to the track with the closure sign and the colleague (56%, 95% CI [55%, 57%]) than to the water reservoir (44%, 95% CI [43%, 45%]). These proportions were significantly different from 50%, $\chi^2(1) = 226.62$, $p < .001$.

We conducted a logistic mixed effects model using the `afex` package (Singmann et al., 2022) in `R` to analyze the probability that participants chose the same first-stage choice action as they had chosen in the previous trial (*stay probability*) as a function of outcome (good vs. bad), transition type (rare vs. common), and their interaction. For each participant, the model included random slopes for transition type, outcome, and their interaction.

Model-based learning from a good outcome increases the stay probability after a common transition but reduces it after a rare transition (and vice versa after a bad outcome). Therefore, an interaction of outcome and transition type would be evidence for model-based learning. Model-free learning would imply increasing the probability of actions that lead to a good outcome and reducing the probability of actions that lead to a bad outcome, independent of whether they followed a common or rare transition. Therefore, a main effect of outcome on the probability of repeating the same action would be evidence for model-free learning.

As seen in Figure 2, we found a significant main effect of outcome on stay probability: participants repeated their choice more often after a good outcome than after a bad outcome, $\chi^2(1) = 5.19$, $p = .023$, OR : 1.08, indicating model-free learning based on our preregistered predictions.

Moreover, we found a significant interaction between transition type and outcome, $\chi^2(1) = 7.87$, $p = .005$, OR : 1.09, indicating model-based learning. Specifically, when the outcome resulted from a rare transition, we found no evidence that the stay probability was affected by whether the outcome was good ($M = 0.886$, 95% CI [0.837, 0.921]) or bad ($M = 0.887$, 95% CI [0.839, 0.922]), $OR = 0.99$, $z = -0.13$, $SE = 0.10$, $p = 0.894$. However, when the outcome resulted from a common transition, the stay probability was higher after good outcomes ($M = 0.902$, 95% CI [0.861, 0.932]) than after bad outcomes ($M = 0.870$, 95% CI [0.818, 0.908]),

[2] In the experiment, of those participants who passed the comprehension check questions, 83% passed on their first attempt, and 17% passed on their second attempt.

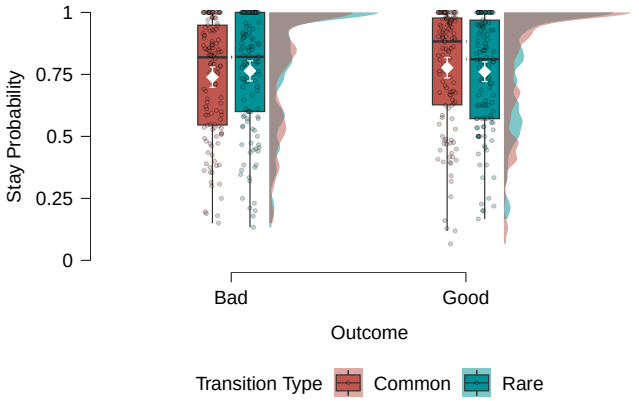

Figure 2: Stay probability was higher for common transitions compared to rare transitions for good outcomes, and lower for common transitions compared to rare transitions for bad outcomes. Mean and 95% CIs are indicated in white. Boxplots indicate median with IQR, and whiskers are constructed using 1.5 times the IQR.

$OR = 1.37$, $z = 4.17$, $SE = 0.10$, $p < .001$.

All results, including exploratory analysis and manipulation check responses, are available on the online repository.

Computational Modeling

To assess model-free and model-based learning on the level of individual participants, we implemented RL models of model-free and model-based moral learning.

Model Specification and Implementation

Model-Free Learning Model-free learning is represented by a simple Q-learning model where people directly learn about the associations between the first-stage choice (direct the wagon towards the reservoir vs. the colleague) and the corresponding second-stage outcome. Specifically, this model learns the anticipated moral values of relying on CBR $Q^{\text{MF}}(\text{CBR})$ versus rules $Q^{\text{MF}}(\text{rules})$. These Q-values are updated based on the prediction error between the anticipated consequences of relying on CBR/rules and the observed outcome. If the choice in trial t was consistent with CBR (i.e., the trolley was directed toward the colleague), the prediction error would be calculated as:

$$\text{MPE}_t = Q_t^{\text{MF}}(\text{CBR}) - R_t, \tag{1}$$

where MPE denotes the moral prediction error, and R denotes the value of the observed consequences. R was defined as 1 when the decision led to good consequences and -1 when it led to bad consequences. The moral prediction error then guides how much the Q-value for reliance on CBR is updated. The size of the update is controlled by the learning rate α:

$$Q_{t+1}^{\text{MF}}(\text{CBR}) = Q_t^{\text{MF}}(\text{CBR}) - \alpha \times \text{MPE}_t, \tag{2}$$

If the participant decided to follow the rule (i.e., the trolley was directed towards the reservoir), the same updating rules are applied to $Q^{\text{MF}}(\text{rules})$.

Model-Based Learning Either action can lead to one of two states: either the trolley will head toward the reservoir ($s_{\text{reservoir}}$) or the trolley will head toward the colleague ($s_{\text{colleague}}$). We modeled model-based learning as people using a model of how likely either decision strategy is to lead to each state, people learning the expected outcomes of both states ($V_t(s_{\text{reservoir}})$ and $V_t(s_{\text{colleague}})$), and using their mental model and the learned values to reason about whether it is better to rely on CBR versus rules. The values of the two states are updated in the same way as the model-free Q-values using Equations 1 and 2, except that they assign value to the states rather than decision strategies. Assuming that people's mental model captures the stated transition probabilities (as in Daw et al. (2011) and a reasonable assumption given the practice trials), we can model the estimates people reach by reasoning about the consequences of relying on CBR versus rules as:

$$\begin{aligned} Q_t^{\text{MB}}(\text{CBR}) &= 0.7 \times V_t(s_{\text{colleague}}) + 0.3 \times V_t(s_{\text{reservoir}}), \\ Q_t^{\text{MB}}(\text{rules}) &= 0.3 \times V_t(s_{\text{colleague}}) + 0.7 \times V_t(s_{\text{reservoir}}). \end{aligned} \tag{3}$$

Decision Execution To implement the assumption that the participants may rely on both model-free and model-based learning and that the weight given to each of the learning mechanisms varies between participants, our model's decisions follow a weighted average of the model-based and model-free Q-values:

$$\begin{aligned} Q_t^{\text{total}}(\text{CBR}) &= w \times Q_t^{\text{MB}}(\text{CBR}) + (1-w) \times Q_t^{\text{MF}}(\text{CBR}), \\ Q_t^{\text{total}}(\text{rules}) &= w \times Q_t^{\text{MB}}(\text{rules}) + (1-w) \times Q_t^{\text{MF}}(\text{rules}), \end{aligned} \tag{4}$$

where w denotes the decision weight of the model-based system, which is estimated for each participant and allows us to quantify how much participants rely on model-free versus model-based learning.

Finally, the decision mechanism is selected probabilistically according to the softmax function:

$$p_t(\text{CBR}) = \frac{e^{\tau \times Q_t^{\text{total}}(\text{CBR})}}{e^{\tau \times Q_t^{\text{total}}(\text{CBR})} + e^{\tau \times Q_t^{\text{total}}(\text{rules})}} \tag{5}$$

Prior Distributions As a prior on the learning rate α, we use a uniform distribution on the interval $[0,1]$. This prior reflects the belief that learning rates > 1 (which would imply changing one's belief by more than the prediction error) and learning rates < 0 (which would imply learning in the opposite direction of the prediction error) are impossible. We also use a uniform distribution on the interval $[0,1]$ as prior on the mixing weight w, which indicates that all weights for model-free versus model-based learning are considered equally likely a priori. As the prior distribution on the temperature parameter τ, we use the lognormal distribution lognormal$(0, 1.4)$, which assigns 90% of the prior probability mass to values of τ between $\frac{1}{10}$ and 10. Finally, we use a prior of Normal(0,1) for the initial values of V_t and Q^{MF}.

Model Implementation We implemented the models in `stan` version 2.35.00 and fitted them using `cmdstanr` version 0.8.0.9000. The model was fitted individually to each participant using four MCMC chains and 20000 iterations per chain, 10000 of which were warm-up iterations.

Prior Predictives & Recovery Simulations

Prior predictive checks confirmed that simulating from the model with different mixture weights w given to model-based versus model-free learning recreates the main patterns in the data associated with these learning types (i.e., the main effect of reward, the interaction of transition type and reward, or both). Further, we conducted recovery simulations where we simulated from the model using different mixture weights w and then fitted the model to the simulated data. This model recovered the weights well, indicating that the mixture parameter can, in principle, be estimated from the data.

Results

Descriptive Results We removed two participants with $\hat{R} > 1.01$ (Vehtari et al., 2021). For all other participants, the model showed good convergence. Overall, the computational modeling results indicate a weak preference for model-free over model-based learning: The average weight of model-based learning was 45%. 57% of participants had a mixture weight w of less than .5 and 26% less than .25 (w ranges from 0 to 1, whereby a value of 0 would indicate only model-free learning took place, and a value of 1 only model-based learning). In contrast, 43% of participants had a w larger than .5 and only 10% larger than .75.

Relationship Between Evidence for Model-Based Learning and Exploratory Measures Because there was low colinearity between the various self-report measures (all VIF $<$ 1.5), we tested the relationship between these variables and evidence for model-based learning (i.e., the posterior median estimate of w) within a single regression model. This indicated no evidence that any of those scales were associated with evidence for model-based learning (all $p > .05$; see online repository for a complete report of all results).

Discussion

In this article, we developed a novel paradigm – the moral dilemma two-step task. Our results from behavioral data and computational modeling provided converging evidence that the paradigm dissociates between model-free versus model-based reinforcement learning in the moral domain. Our results corroborate that the arbitration between different mechanisms of moral decision-making is shaped by learning from the consequences of previous decisions (Maier et al., 2024). We found that this learning combines model-free and model-based learning. Although the weights given to each mechanism differ considerably between participants, most, if not all, relied on both mechanisms to some extent.

One alternative explanation for our findings could be that rather than learning about decision mechanisms, people learned about specific behaviors (e.g., whether to press F or J; whether to direct a wagon toward a colleague or a water reservoir). Based on the evidence for metacognitive moral learning (Maier et al., 2024), which generalized across different dilemmas and moral decision-making tasks, it is unlikely that this is the only type of learning that occurred in our experiment. However, purely behavioral learning cannot be entirely ruled out with the data collected in the current version of the moral dilemma two-step task. We are currently working on follow-up experiments that directly delineate metacognitive and behavioral learning within the two-step task by using different decision scenarios in different trials (to test for transfer between them) and adding transfer measures after the experiment.

Another interesting avenue for future work is to investigate which traits explain how much people rely on model-based versus model-free learning. In the current study, we did not find a relationship between different exploratory scales and individual-level evidence for model-based versus model-free learning. One explanation is that we calibrated the statistical power to be able to detect model-based and model-free learning in the choice data rather than for relating evidence for either learning type to questionnaire measures. Given the low reliability of tasks as individual difference measures (cf. Pedroni et al., 2017; Schuch et al., 2022), the study likely had much lower power for the latter purpose. We are planning follow-up studies to test these relationships with increased sample size and number of trials to allow for a more powerful test of determinants of model-based versus model-free moral learning. Further, our computational modeling approach made a simplifying assumption by using only two reward values for outcomes (one for positive and one for negative outcomes). Future work should measure how participants experience the outcomes and use this information to model the prediction error more accurately.

The moral dilemma two-step task allows researchers to study and dissociate model-free and model-based moral learning from consequences, making it a valuable tool for research on moral learning from outcomes. Further, by modifying the learning signal our paradigm could straightforwardly be adapted to investigate other types of moral learning, such as learning from social feedback, creating applicability in a variety of research areas in moral psychology and beyond.

At a broader level, understanding the mechanisms underlying metacognitive moral learning informs our understanding of moral development and moral progress. This improved understanding may, in turn, strengthen the scientific foundations for fostering moral development, and help inform researchers and educators in developing interventions that empower individuals to translate their ethical commitments into action.

References

Alexander, L., & Moore, M. (2021). Deontological ethics (E. N. Zalta, Ed.). `https://plato.stanford`

.edu / archives / win2021 / entries / ethics -deontological/

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, *5*(2), 187–202.

Bentham, J. (1789). *An introduction to the principles of morals and legislation*.

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366.

Crone, D. L., & Laham, S. M. (2017). Utilitarian preferences or action preferences? de-confounding action and moral code in sacrificial dilemmas. *Personality and Individual Differences*, *104*, 476–481.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https : / / doi .org/10.1038/nn1560

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The cni model of moral decision-making. *Journal of Personality and Social Psychology*, *113*(3), 343–376.

Gigerenzer, G. (2008). Moral intuition = fast and frugal heuristics? In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 1–26). MIT Press.

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct? *Emotion*, *16*(8), 1107.

Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164.

Kant, I., & Schneewind, J. B. ([1785] 2002). *Groundwork for the metaphysics of morals*. Yale University Press.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off? *PLoS computational biology*, *12*(8), e1005090.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762.

Lockwood, P. L., Klein-Flügge, M. C., Abdurahman, A., & Crockett, M. J. (2020). Model-free decision making is prioritized when learning to avoid harming others. *Proceedings of the National Academy of Sciences*, *117*(44), 27719–27730. https : / / doi .org / 10 .1073/pnas.2010890117

Maier, M., Cheung, V., & Lieder, F. (2024). Metacognitive learning from consequences of past choices shapes moral decision-making. https : / / doi .org / 10 .31234/osf.io/gjf3h

Mata, A., Vaz, A., & Mendonça, B. (2022). Deliberate ignorance in moral dilemmas: Protecting judgment from conflicting information. *Journal of Economic Psychology*, *90*, 102523.

Mill, J. S. (1879). *Utilitarianism*. Fraser's Magazine.

Patil, I., Zucchelli, M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2020). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology*, *120*. https : //doi.org/10.1037/pspp0000281

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, *1*(11), 803–809.

Schuch, S., Philipp, A. M., Maulitz, L., & Koch, I. (2022). On the reliability of behavioral measures of cognitive control: Retest reliability of task-inhibition effect, task-preparation effect, stroop-like interference, and conflict adaptation effect. *Psychological research*, *86*(7), 2158–2184. https : //doi .org/10 .1007/ s00426-021-01627-x

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2022). *Afex: Analysis of factorial experiments* [R package version 1.1-1]. https : // CRAN.R-project.org/package=afex

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc. *Bayesian analysis*, *1*(1), 1–28. https://doi.org/10.1214/20-BA1221

Williams, E. G. (2023). Rule utilitarianism and rational acceptance. *The Journal of Ethics*, 1–24. https : // doi.org/10.1007/s10892-023-09428-7