Every Pull a Lesson

Moral history in a multi-turn trolley game

Veit Heller

cyberwitchery lab

August 2026

lab note · version 1.0

abstract

Would your trolley choice change if you knew what the person on the side track had chosen when they previously held the lever? This note turns that question into a game. People rotate between the lever and the tracks. Their choices are remembered. A person who judges others today may be judged by them tomorrow.

One extra turn is enough to create a strategic link. It also makes later treatment coarse in a way one-to-one reciprocity models do not: rewarding or punishing one person means choosing the fate of everyone who shares their track. A private grudge can therefore create public collateral harm and new grudges.

We study the two-turn case, derive a threshold for collateral retaliation, and compare simple moral rules in mixed populations. Fixed rule mixtures remain solvable when the former actor can occupy either track. We finish by specifying a finite-agent model for the harder case, where histories and relationships evolve together.

keywords trolley problem; repeated games; moral history; reciprocity; agent-based models

the idea

The trolley becomes a game when a choice follows its author across positions.

1 One extra turn

In the familiar switch case, a trolley will strike several people unless a bystander diverts it onto a side track where it will strike one. The case asks whether intervening to harm one is better than allowing greater harm (Foot 1967; Thomson 1976). It is deliberately timeless. Nobody has a history, and nobody must live with another person's memory of the choice.

Now let the roles rotate. Someone who controls the lever in one turn may occupy a track in the next. The new decision-maker can see what that person did when the roles were reversed. Our starting question is:

> Would your choice change if you knew the choice previously made by the person on the side track?

The answer could change for several reasons. The earlier act may serve as a precedent. It may reveal character. It may create gratitude or resentment. A later mover may also refuse to use present harm as a reward or punishment for past conduct. Once the first mover expects any of these responses, the original trolley choice becomes strategic as well as moral.

No intervention fee or helping payoff is needed. The only material event is still being struck by the trolley. The change is that today's choice can alter tomorrow's route.

1.1 What is and is not new

The strategic loop is familiar. In image-scoring models, a donation changes the donor's reputation and therefore the donations they later receive (Nowak and Sigmund 1998; Wedekind and Milinski 2000; Engelmann and Fischbacher 2009). Our game has the same skeleton: an actor becomes a recipient and anticipates how a recorded act will affect later treatment. Remembered conduct creating incentives is the starting point, but importantly not yet a novelty claim.

Other trolley work supplies separate pieces. Iterative sacrificial dilemmas give targets a history of what happened to them (Bostyn and Roets 2022), while our players also carry a history of what they did. Everett et al. let a trolley judgment affect later trust through a transfer in another game (Everett et al. 2016). Ethical dilemmas have also been formalized as strategic games (Naumov and Yew 2021), and MoralityGym provides sequential trolley-style environments for computational agents (Rosen et al. 2026).

Dickinson's Payoff-Trolley allocates real monetary losses among anonymous track occupants, but its safe decision-maker never becomes a later target (Dickinson 2026). In Loriga's Looping Trolley, pulling passes control to a new actor and doubles the future victims; earlier actors remain observers rather than returning to the tracks (Loriga 2026).

The difference here lies in the stage game. The mover bears no current material loss and chooses which of two sets of other players is struck. Both actions allocate harm. Once history singles out one occupant for reward or punishment, the track bundles that person with bystanders. Work on collective punishment studies harm to innocents (Pereira and van Prooijen 2018), here the bundling is mechanical and randomized. The contribution of this note is to work out the consequences of that coarse instrument and connect it to collaterality.

2 The game

2.1 One turn

There are $m + 2$ active roles. One person controls the lever. One person, s, is alone on the side track, we call this person the singleton. A set M of $m \geq 2$ people occupies the main track. The trolley initially heads toward M. The mover chooses $a \in \{0, 1\}$:

$$a = 1 \quad \text{means pull toward } s, \qquad a = 0 \quad \text{means stay on } M.$$

Let $L > 0$ be the loss from being struck. The material payoff to player i in that turn is

$$\pi_i(a) = -L\,\mathbf{1}\{i \text{ is on the struck track}\}. \tag{1}$$

The mover is safe in the current turn, so their material payoff is zero after either action. This is part of the trolley problem, not a missing incentive, and the game works without it. Let Δ_i denote mover i's present moral reason for pulling rather than staying. It may represent concern for casualties, reluctance to intervene, a rule, or something else. In particular, it can represent a difference between harm caused by acting and harm allowed by omission (Spranca et al. 1991). In a one-turn game, i pulls when $\Delta_i \geq 0$.

Importantly, no one exits the game after being struck.

2.2 The smallest loop

The smallest multi-turn game has two turns. In turn 1, A controls the lever and chooses a_A. In turn 2, A is the singleton, B controls the lever, and B knows what A chose. Write

$$p_x = \Pr(B \text{ pulls in turn 2} \mid a_A = x), \qquad x \in \{0, 1\}.$$

These probabilities can describe a mixed strategy, uncertainty about B's rule, or a population of possible later movers.

If B pulls, A is struck and one person is harmed. If B stays, A is spared and m other people are struck. Define

$$\chi_A = U_A(B \text{ stays}) - U_A(B \text{ pulls}).$$

A positive χ_A means that A prefers to be spared, a negative value means that concern for the other casualties is stronger. Under pure self-preservation, $\chi_A = L$. If, in addition, A assigns moral cost $\alpha_A L$ to each casualty, counting themselves among the casualties, then

$$\chi_A = L - \alpha_A L(m - 1).$$

This convention gives A both the material loss L and moral concern for themselves when struck. If moral concern counts only other people, the same calculation gives $\chi_A = L - \alpha_A L m$ instead.

The same utility can generate the present and future terms. For example, suppose A values their own material payoff and assigns moral cost $\alpha_A L$ to every casualty in both turns. Because A is safe while moving, their first-turn moral reason is $\Delta_A = \alpha_A L(m-1)$. With the self-counting convention above, their continuation comparison is $\chi_A = L - \alpha_A L(m-1)$. Thus Δ_A and χ_A are two comparisons made by one utility, not two costs added to the same event. We keep Δ_A abstract so that it can also include act-omission aversion or a rule-based reason.

With continuation weight $\delta \in (0, 1]$, A's utility difference between pulling and staying in turn 1 is

$$U_A(1) - U_A(0) = \Delta_A - \delta\chi_A(p_1 - p_0). \tag{2}$$

The first term is the ordinary trolley judgment. The second is the effect of the moral history that the choice creates.

Proposition 1 (When the first turn becomes strategic). *Later treatment adds the term $-\delta\chi_A(p_1 - p_0)$ to the utility difference between A's first-turn actions. It vanishes when later movers ignore the earlier action, $p_1 = p_0$, or when A is indifferent between the two continuation outcomes, $\chi_A = 0$.*

When $\chi_A > 0$, the term favors pulling if pullers are later protected ($p_1 < p_0$) and staying if pullers are later punished ($p_1 > p_0$). When $\chi_A < 0$, those directions reverse.

Proof. Subtract the expected continuation utility after $a_A = 0$ from the expected continuation utility after $a_A = 1$. The difference is $-\delta\chi_A(p_1 - p_0)$. Its zeros and sign give the result. □

The proposition does not tell us which action is rewarded. That depends on the later mover's moral rule. The proposition captures a simple point: history matters strategically exactly through the treatment it changes and the value the earlier mover places on that treatment.

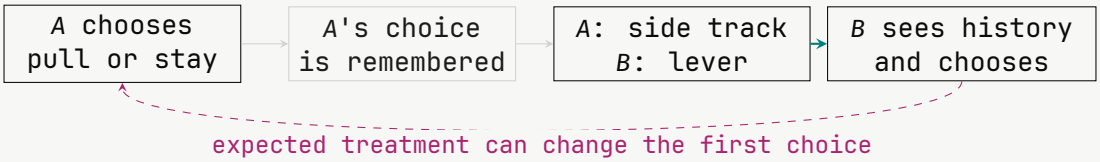

Figure 1. The two-turn loop. The first action travels forward as history; the expected response travels backward as an incentive.

2.3 Longer games

In a longer game, roles keep rotating and actions accumulate in a history h. Let $V_i(h, a)$ be player i's continuation value after current action a. The mover pulls when

$$\Delta_i(h) + \delta\big[V_i(h, 1) - V_i(h, 0)\big] \geq 0. \tag{3}$$

This is the same idea as Equation (2), with all later encounters folded into V_i. The hard part is no longer the inequality, but rather shifts into describing the history, the rule used to judge it, and the way relationships change after each route choice.

In plainer words, it becomes a true moral decision framework with encoded relationships between the actors.

3 Relationships and collateral harm

A mover may care not only about how many people are struck, but about who they are. Let $H(a)$ be the set struck by action a, so $H(1) = \{s\}$ and $H(0) = M$. For mover i, let $\alpha_i \geq 0$ be the weight placed on each casualty. Let w_{ij} be i's directed regard for player j. Positive w_{ij} means that i especially wants to protect j; negative w_{ij} means hostility. Use the simple decision utility

$$u_i(a \mid w_i) = -L\left[\alpha_i |H(a)| + \sum_{j \in H(a)} w_{ij}\right]. \tag{4}$$

This is a moral score for the decision, not the mover's current material payoff. Pulling is chosen exactly when

$$\alpha_i(m - 1) + \sum_{j \in M} w_{ij} - w_{is} \geq 0. \tag{5}$$

The first term favors fewer casualties. The remaining terms say that identities can outweigh that numerical advantage.

This also connects the relationship model to the dynamic notation in Equation (3). If $\mathbf{w}_i = \mathbf{w}_i(h)$ records the residue of the history, the present moral reason is

$$\Delta_i(h) = \Delta_i(\mathbf{w}_i(h)) = L\left[\alpha_i(m - 1) + \sum_{j \in M} w_{ij} - w_{is}\right].$$

The bracket is the left side of Equation (5). A forward-looking player adds the continuation term from Equation (3); a myopic player uses the bracket alone.

Proposition 2 (The price of favor and vengeance)**.** *Suppose every current track occupant is neutral to mover i except target j.*

If $w_{ij} = -v < 0$, vengeance agrees with casualty minimization when j is the singleton. When j is on the main track, vengeance makes i stay if and only if

$$v > \alpha_i(m - 1).$$

If $w_{ij} = g > 0$, favor agrees with casualty minimization when j is on the main track. When j is the singleton, favor makes i stay if and only if $g > \alpha_i(m - 1)$. Each conflicting choice raises the number struck from one to m.

Proof. For a hostile singleton, the left side of Equation (5) is $\alpha_i(m - 1) + v$, so pulling both harms the target and minimizes casualties. For a hostile main-track target it is $\alpha_i(m - 1) - v$, which is negative exactly above the stated threshold. Replacing hostility $-v$ with favor g switches the two positions. □

The main-track case contains the distinctive moral cost. "You hit me, so I hit you" cannot be carried out against one person anymore. Staying also strikes that person's $m - 1$ trackmates. The grudge is dyadic, but the action is not.

motive	target position	route serving the motive	result
harm	side	pull	target alone struck
harm	main	stay	target and $m - 1$ others struck
protect	side	stay	m other people struck
protect	main	pull	target spared with co-occupants

Table 1. Track position decides whether a personal motive agrees with reducing casualties. Each conflict adds $m - 1$ casualties relative to pulling.

3.1 How a grudge spreads

Relationships can change after each turn. Let $H_t(a)$ be the struck track occupants and $S_t(a)$ the spared track occupants. Neither set includes the mover. Let d_t be the mover. One simple update is

$$w_{ij,t+1} = \rho w_{ij,t} + \mathbf{1}\{j = d_t\}\left[g_i \mathbf{1}\{i \in S_t(a_t)\} - v_i \mathbf{1}\{i \in H_t(a_t)\}\right], \tag{6}$$

where $\rho \in [0, 1]$ controls memory. Being spared creates gratitude g_i; being struck creates resentment v_i. This version credits or blames only the mover. A larger model could also let a collateral victim blame the friend whom the mover protected.

Staying to strike an enemy on the main track also strikes $m - 1$ co-occupants, who may now resent the mover. A larger m creates more possible branches but also raises the threshold $\alpha_i(m - 1)$ in Proposition 2. Under uniform assignment, a target known to be on a track is on the main track with probability $m/(m + 1)$. Whether a grievance survives until the relevant people meet again depends on ρ, grievance strength, and the assignment process. These

quantities are generated by the finite game, so Section 5 treats the cascade rate as an output.

Retaliation and counter-retaliation are familiar in repeated games (Nikiforakis 2008; Nikiforakis et al. 2012; Bolle et al. 2014). Here one attempt to settle a score can give several uninvolved people new scores to settle. This resembles collective punishment, but the bystanders are hit because the instrument cannot separate track-mates (Pereira and van Prooijen 2018).

4 Rules in mixed populations

We now ask how simple rules fare when they repeatedly judge one another. These are operational rules, not complete moral philosophies. The point is to compare their consequences inside this game.

The full one-period state relevant to a former actor contains two facts: their previous action $x \in \{0, 1\}$ and their current position $z \in \{\text{side}, \text{main}\}$. A deterministic response rule is therefore

$$\tilde{f} : \{0, 1\} \times \{\text{side}, \text{main}\} \longrightarrow \{0, 1\}.$$

This representation isolates one focal former actor. If several current track occupants' histories matter at once, the state and rule domain are larger. There are sixteen rules in the focal-actor slice. Four readable examples are shown in Table 2.

	former actor on side		former actor on main		
rule	stay	pull	stay	pull	**reading**
C	1	1	1	1	minimize current casualties
O	0	0	0	0	never intervene
I	0	1	0	1	repeat the earlier action
D	1	0	0	1	punish stayers, protect pullers

Table 2. Four position-aware rules. Entries give the current action, with 1 = pull and 0 = stay. The other twelve deterministic rules are hybrids.

Track position changes the meaning of these rules. When the former actor is the singleton, *I* and *D* are exact opposites. In that position, *I* can also be read as rewarding stayers, while *D* can be read as anti-imitation. When the former actor is on the main track, the two maps coincide: repeating a stay punishes the stayer, and repeating a pull protects the puller. Behavior in one position therefore does not identify the motive behind the rule. Observing both positions tells us more. Small deterministic norms have a close precedent in work on indirect reciprocity (Ohtsuki and Iwasa 2006). The trolley merely adds track position to the rule's input.

4.1 A position-aware benchmark

The sixteen-rule system is still tractable if placement is independent. Let μ_k be the population share of rule k. After each choice, place its author on the side track with probability σ and on the main track with probability $1 - \sigma$. Draw the next mover independently from the population. Uniform placement among the $m + 1$ track slots gives $\sigma = 1/(m + 1)$. For $z \in \{\mathrm{S}, \mathrm{M}\}$, define

$$F_{xz} = \sum_k \mu_k \tilde{f}_k(x, z), \qquad \overline{F}_x = \sigma F_{x\mathrm{S}} + (1 - \sigma) F_{x\mathrm{M}}. \tag{7}$$

Thus F_{xz} is the pull rate after action x when its author now occupies position z, and $\overline{F}_x$ averages over that position. Except when $\overline{F}_0 = 0$ and $\overline{F}_1 = 1$, the stationary pull rate is

$$q^* = \frac{\overline{F}_0}{1 - \overline{F}_1 + \overline{F}_0}. \tag{8}$$

Write $\pi_0 = 1 - q^*$ and $\pi_1 = q^*$. A type-k mover pulls at stationary rate

$$q_k = \sum_{x \in \{0,1\}} \pi_x \left[\sigma \tilde{f}_k(x, \mathrm{S}) + (1 - \sigma) \tilde{f}_k(x, \mathrm{M})\right]. \tag{9}$$

If that mover chose a, their probability of being struck in the next turn is

$$e_a = \sigma F_{a\mathrm{S}} + (1 - \sigma)(1 - F_{a\mathrm{M}}).$$

Their next-turn exposure is therefore

$$h_k = (1 - q_k) e_0 + q_k e_1. \tag{10}$$

Two aggregate quantities follow from the same calculation. The expected number struck per turn is

$$\bar{d}(\mu) = q^* + m(1 - q^*) = m - (m - 1) q^*, \tag{11}$$

and the expected number of bystanders struck alongside a focal former actor on the main track is

$$b(\mu, \sigma) = (m - 1)(1 - \sigma) \sum_{x \in \{0,1\}} \pi_x (1 - F_{x\mathrm{M}}). \tag{12}$$

Proposition 3 (Position-aware rule mixtures). *Under independent type draws and independent placement with side probability σ, Equations* (8) *to* (12) *give the stationary pull rate, each type's action rate and next-turn exposure, total harm, and bundled bystander harm for any mixture of the sixteen focal-actor rules.*

Proof. Averaging first over mover type and then over the former actor's position gives the two-state transition probabilities $\overline{F}_x$ and hence Equation (8). Averaging rule k over the stationary previous action and position gives Equation (9). A former mover who chose a is struck if the next mover pulls while they are the singleton, or stays while they are on the main track. This gives e_a and Equation (10). Each pull strikes one and each stay strikes m, giving Equation (11). Finally, a main-track stay strikes the focal former actor together with $m - 1$ co-occupants, which gives Equation (12). □

When $\overline{F}_0 = 1$ and $\overline{F}_1 = 0$, the stationary distribution in Equation (8) is one half, but a chain started from a fixed action alternates forever. In that periodic case, one half is also the long-run time average. When $\overline{F}_0 = 0$ and $\overline{F}_1 = 1$, both actions are absorbing and the initial condition selects the limit.

The pure D population makes the bundled harm concrete. For $0 < \sigma < 1$, its stationary pull rate, type action rate, and next-turn exposure are all one half:

$$q^* = q_D = h_D = \frac{1}{2}, \qquad b(\mu, \sigma) = \frac{(m-1)(1-\sigma)}{2}.$$

A former stayer is struck in either position, while a former puller is spared. In half of the turns with the former actor on the main track, the rule stays to punish a former stayer and strikes $m - 1$ bystanders. Under uniform track placement, bundled bystander harm is $m(m-1)/[2(m+1)]$ per turn.

4.2 The singleton slice

For a first closed-form comparison, fix the former actor as the singleton. A rule then reduces to a map $f : \{0,1\} \to \{0,1\}$. The four restrictions are

$$C = (1,1), \qquad O = (0,0), \qquad I = (0,1), \qquad R = (1,0).$$

Restricting Table 2 to the singleton columns gives exactly these four maps; R is the singleton form of D. They exhaust this restricted strategy space. As the preceding discussion suggests, their labels describe behavior rather than uniquely identifying a motive.

Let μ_k be the population share of rule k and define

$$F_x = \sum_k \mu_k f_k(x).$$

This is $\sigma = 1$ in Proposition 3, so $\overline{F}_x = F_x$, $e_a = F_a$, and

$$h_k = (1 - q^*)F_{f_k(0)} + q^* F_{f_k(1)}.$$

Here h_k is exposure in the forced next-turn slot. Under type-independent assignment, other track exposure is the same for every type, so differences in h_k also give the differences in total expected material loss. This need not hold when role assignment depends on type.

These formulas are exact with independent sampling. In a fixed finite group, the previous mover is unavailable to move next because they are the singleton. The next type is drawn from the remaining people, so exact calculations need a finite-state transition matrix. The difference can be large in small groups.

The same formulas let us connect this population model to the strategic term in Equation (2). Suppose a purely self-interested mover knows that they will be the next singleton. Pulling now gives expected continuation loss δLF_1; staying gives δLF_0. More generally, a mover with present moral reason Δ_i pulls exactly when

$$\Delta_i \geq \delta L(F_1 - F_0). \tag{13}$$

Proposition 4 (Material best response to a rule population). *For a purely self-interested, one-step-ahead mover, the best response at every observed history is unconditional pulling when $F_1 < F_0$ and unconditional staying when $F_0 < F_1$. If $F_0 = F_1$, the mover is materially indifferent.*

When $F_0 \neq F_1$, the matching unconditional rule weakly dominates each conditional rule as a contingent plan. The comparison is strict on average whenever both histories occur with positive probability.

Proof. The mover is safe now and is struck next turn exactly when the next mover pulls. Their material continuation payoff from current action a is therefore $-\delta LF_a$, which does not depend on the history they observed. Choosing the smaller of F_0 and F_1 at both histories gives the result. □

This proposition does not imply that conditional moral rules are mistakes; it shows that they are not generated by material self-protection alone in this special case. Rules *I* and *R* encode a preference about precedent or desert. Their users may accept a greater chance of being struck in order to apply that rule.

It also explains why strategic anticipation disappears in a *C*–*O* population: there $F_0 = F_1$, so current conduct does not change later risk. The one-step qualifier matters. A fully forward-looking mover can also care about how the current action changes later states of the public action chain and hence later main-track exposure. Whether the unconditional best response survives that full-horizon problem is left open here.

Proposition 5 (Rule mixtures in the singleton slice). *The following comparisons hold.*

1. *With share x of C and $1 - x$ of O, $q^* = x$ and $h_C = h_O = x$.*
2. *With share $r \in (0, 1)$ of R and $1 - r$ of O,*

$$q^* = \frac{r}{1 + r}, \qquad h_O = r, \qquad h_R = \frac{r^2}{1 + r} < h_O.$$

3. *With share $x \in (0, 1)$ of C and $1 - x$ of R,*

$$q^* = \frac{1}{2 - x}, \qquad h_C = x, \qquad h_R = \frac{1 + x - x^2}{2 - x} > h_C.$$

4. *A pure I population preserves its initial action. Adding any positive share of C eventually fixes pulling, adding any positive share of O eventually fixes staying. In an I–R mixture with both shares positive,* $q^* = h_I = h_R = 1/2$. *A pure R population instead alternates, with stationary distribution and long-run time average* $1/2$.
5. *With positive shares* c, r, o *of* C, R, O *summing to one,*

$$q^* = \frac{c + r}{1 + r}, \qquad h_C = c, \qquad h_R = c + \frac{r(c + r)}{1 + r}, \qquad h_O = c + r.$$

Hence $h_C < h_R < h_O$. *Holding* c *fixed and replacing* O *with* R *raises pulling at rate* $\partial q^*/\partial r = (1 - c)/(1 + r)^2 > 0$.

Proof. For C–O, $F_0 = F_1 = x$. For R–O, $F_0 = r$ and $F_1 = 0$. For C–R, $F_0 = 1$ and $F_1 = x$. Substitution into Equations (8) and (10) gives the first three results, including

$$h_R - h_C = \frac{1 - x}{2 - x} > 0.$$

For pure I, $F_0 = 0$ and $F_1 = 1$, so both actions are absorbing. Adding C makes only pulling absorbing, adding O makes only staying absorbing. In an I–R mixture with share $r \in (0, 1)$ of R, $F_0 = r$ and $F_1 = 1 - r$, which gives $q^* = h_I = h_R = 1/2$. At $r = 1$, the transition is deterministic reversal, so the action alternates.

Finally, in a C–R–O mixture, $F_0 = c + r$ and $F_1 = c$. Substitution gives the fifth set of formulas. Moreover,

$$h_R - h_C = \frac{r(c + r)}{1 + r} > 0, \qquad h_O - h_R = \frac{r(1 - c)}{1 + r} > 0.$$

Differentiating $q^* = (c + r)/(1 + r)$ while holding c fixed gives the stated rate. □

The first comparison separates aggregate harm from personal exposure. More unconditional pullers reduce total harm, but C and O users remain equally likely to be struck. Conditional treatment changes that exposure. In an R–O mixture, R users are struck less often; in a C–R mixture, C users are struck less often.

The three-type comparison puts both effects together: replacing O with R raises pulling, while R users remain less protected than C users who pull without enforcing. This resembles second-order free-riding in costly-punishment models (Ohtsuki, Iwasa, and Nowak 2009), although here the cost is exposure to later harm rather than a fee paid to punish.

The singleton slice leaves out collateral branching because punishment strikes the former actor alone. The position-aware benchmark restores main-track punishment, bundled bystander harm, and the other twelve deterministic rules. What it does not restore is an evolving relationship network. The rule still conditions on one focal actor and one remembered action.

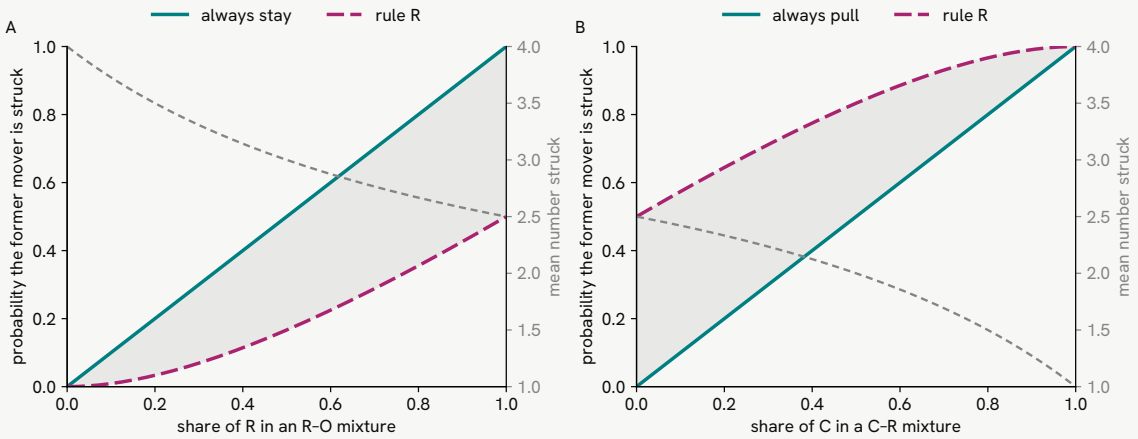

Figure 2. Material performance when the former mover becomes the singleton, for $m = 4$. Panel A mixes R with O, Panel B mixes C with R. Colored lines show the probability that a former mover is struck. Gray dashed lines show the expected number struck per turn. At the pure-R endpoint, the values are stationary and long-run time averages of the alternating process.

5 A computational version

The paper's supplementary materials include a computational simulation described in this section.

Fixed rule mixtures under independent placement are solved in Proposition 3. Simulation becomes useful when the group is finite, several occupants' histories matter at once, or directed relationships evolve. A finite-agent model can keep those additions transparent.

5.1 State and turn order

Consider N agents. At turn t, the state is

$$X_t = (h_t, w_t, \tau_t, \ell_t).$$

Here h_t is the history retained by the game. At minimum it records the previous action and mover; richer versions can retain positions, outcomes, or the complete sequence. The matrix w_t records directed relationships, τ_t records each agent's rule or moral parameters, and ℓ_t records cumulative losses. The initial history h_0 must include a seed action or a rule for generating one. This choice matters especially in populations of precedent followers. A turn has five steps:

1. Assign one mover, one singleton, and m main-track occupants.
2. Show the mover the current placement and the histories allowed by the information rule.
3. Choose pull or stay.
4. Add L to the losses of everyone on the struck track.
5. Update directed regard with Equation (6).

Role assignment is part of the model. Independent sampling from a large population gives a mean-field benchmark. Random assignment without replacement within a fixed group gives an exact small-group game, while deterministic rotation is a different scheme. We can also force the previous mover onto a chosen track when we want to isolate a particular reciprocal encounter. That forced placement is the baseline for comparing the sixteen one-period rules: the previous mover is the focal historical actor, while the other track occupants are bystanders.

5.2 Two kinds of agents

A first computational pass can compare two agent designs.

Rule agents. Each agent carries one of the sixteen maps $\tilde{f}(x, z)$. These agents make it easy to enumerate mixtures, find absorbing conventions, and check which rules protect their users.

Relationship agents. Each agent carries parameters $\alpha_i, g_i, v_i, \rho_i$ and chooses using Equation (5). Their directed weights evolve after every turn. These agents can forgive, form grudges, and treat two people with identical public records differently.

A later version might add foresight. A relationship agent then compares the two current moral scores plus a simulated continuation value, as in Equation (3). Short Monte Carlo rollouts are enough to ask whether an agent changes its present action because it expects to occupy a track later. We leave this modelling open for future work.

5.3 What to vary

The smallest useful parameter sweep varies:

- main-track size m,
- memory ρ,
- the distributions of α, g, and v,
- the mixture of deterministic rules,
- finite group size N and the role-assignment process, and
- whether histories contain only actions or also identities, outcomes, and track positions.

Useful outputs are plain: pulls per turn, people struck per turn, loss by rule, collateral losses outside the intended relationship, number and duration of grudges, and the size of the largest grievance cascade.

For cascades, define an empirical local branching ratio R_C as the mean number of new collateral grievances that later produce a collateral retaliation, per collateral retaliation, measured while the grievance network is still sparse. This is an output of the specified assignment, memory, and preference process. It is not an additional behavioral parameter or a claim about the eventual size of a feud. The direct target's response is excluded because it is not a collateral branch.

In a static comparison, agent types remain fixed. Any claim about one rule spreading or disappearing requires an additional population-update rule, such as payoff-weighted imitation with mutation. That choice should be part of the model rather than smuggled into the language of "fitness."

5.4 Checks before exploration

The code should first reproduce the results we already know:

1. In the independently sampled focal-actor model, recover Proposition 3, including bundled bystander harm.
2. In the singleton slice, recover the history-independent best response in Proposition 4.
3. In the same model, recover Proposition 5.
4. With $w_{ij} = 0$, always choose the casualty-minimizing route when $\alpha_i > 0$.
5. With one nonzero relationship, recover the threshold in Proposition 2.
6. Near an empty grievance network, estimate R_C directly and record how early cascade generations behave on either side of one.

Only then should we use the simulation for questions without closed forms, such as whether collateral vengeance survives in finite groups, whether precedent followers lock in the first convention, how an explicit population-update rule changes the mixture, and where the mean-field picture fails.

6 Limits of this version

The closed-form population results use one-period memory, independent type draws, and independent placement of the previous mover. They solve the position-aware focal-actor rule model, including bundled bystander harm, but not a fixed finite group in which several occupants carry relevant histories.

The best-response result is also one-step-ahead. A full-horizon material best response must account for the effect of today's action on later states of the action chain and remains unsolved in this version.

The response probabilities in the two-turn model and the rule shares in the population model are exogenous. We derive a best response to them, but not an equilibrium rule distribution. Nor do differences in the probability of being struck constitute an evolutionary dynamic by themselves. Selection requires a separate account of reproduction, learning, or imitation.

Finally, Equation (6) is one deliberately simple account of moral memory. It assigns credit and blame to the mover, uses geometric decay, and does not let players dispute responsibility. It also treats action and omission symmetrically: a singleton may feel gratitude when the mover stays, even though the trolley was never diverted away from them. An action-sensitive update could give intervention and omission different weights. Other update rules may change the shape or even the existence of grievance cascades.

7 What the model visualizes

The ordinary trolley problem asks what to do at one lever. The multi-turn version asks what happens when the people on the tracks have made the same kind of choice before and may make it again.

Three points organize the note.

First, moral history creates strategy without adding an artificial cost. If later treatment depends on earlier conduct, the first mover has something to anticipate.

Second, imitation and treatment of a person are different ideas. They may even recommend opposite routes. Track position tells us when they coincide and when they come apart.

Third, the trolley is a coarse instrument. A mover cannot punish or protect one person independently of their track-mates. This makes collateral harm part of the game rather than an external side effect. It also lets a dyadic grievance branch into a social process.

None of this tells us that history *should* govern the route. Refusing to turn current harm into punishment may be the most defensible rule, especially when uninvolved people share the target's track. The game is useful because it makes that refusal, and its alternatives, precise enough to compare.

References

Bolle, Friedel, Jonathan H. W. Tan, and Daniel John Zizzo (2014). "Vendettas". In: *American Economic Journal: Microeconomics* 6.2, pp. 93–130. DOI: 10.1257/mic.6.2.93.

Bostyn, Dries H. and Arne Roets (2022). "Sequential Decision-Making Impacts Moral Judgment: How Iterative Dilemmas Can Expand Our Perspective on Sacrificial Harm". In: *Journal of Experimental Social Psychology* 98, p. 104244. DOI: 10.1016/j.jesp.2021.104244.

Dickinson, David L. (2026). *Consequential Ethical Dilemmas: The Payoff-Trolley Game*. Discussion Paper 18428. IZA Institute of Labor Economics.

Engelmann, Dirk and Urs Fischbacher (2009). "Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game". In: *Games and Economic Behavior* 67.2, pp. 399–407. DOI: 10.1016/j.geb.2008.12.006.

Everett, Jim A. C., David A. Pizarro, and Molly J. Crockett (2016). "Inference of Trustworthiness from Intuitive Moral Judgments". In: *Journal of Experimental Psychology: General* 145.6, pp. 772–787. DOI: 10.1037/xge0000165.

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

Loriga, Leandro (2026). *The Looping Trolley and Temporal Extension of Moral Responsibility*. SSRN Working Paper 6899558. Social Science Research Network. DOI: 10.2139/ssrn.6899558.

Naumov, Pavel and Rui-Jie Yew (2021). "Ethical Dilemmas in Strategic Games". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.13, pp. 11613–11621. DOI: 10.1609/aaai.v35i13.17381.

Nikiforakis, Nikos (2008). "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" In: *Journal of Public Economics* 92.1–2, pp. 91–112. DOI: 10.1016/j.jpubeco.2007.04.008.

Nikiforakis, Nikos, Charles N. Noussair, and Tom Wilkening (2012). "Normative Conflict and Feuds: The Limits of Self-Enforcement". In: *Journal of Public Economics* 96.9–10, pp. 797–807. DOI: 10.1016/j.jpubeco.2012.05.014.

Nowak, Martin A. and Karl Sigmund (1998). "Evolution of Indirect Reciprocity by Image Scoring". In: *Nature* 393, pp. 573–577. DOI: 10.1038/31225.

Ohtsuki, Hisashi and Yoh Iwasa (2006). “The Leading Eight: Social Norms That Can Maintain Cooperation by Indirect Reciprocity”. In: *Journal of Theoretical Biology* 239.4, pp. 435–444. DOI: 10.1016/j.jtbi.2005.08.008.

Ohtsuki, Hisashi, Yoh Iwasa, and Martin A. Nowak (2009). “Indirect Reciprocity Provides Only a Narrow Margin of Efficiency for Costly Punishment”. In: *Nature* 457.7225, pp. 79–82. DOI: 10.1038/nature07601.

Pereira, Andrea and Jan-Willem van Prooijen (2018). “Why We Sometimes Punish the Innocent: The Role of Group Entitativity in Collective Punishment”. In: *PLOS ONE* 13.5, e0196852. DOI: 10.1371/journal.pone.0196852.

Rosen, Simon, Siddarth Singh, Ebenezer Gelo, Helen Sarah Robertson, Ibrahim Suder, Victoria Williams, Benjamin Rosman, Geraud Nangue Tasse, and Steven James (2026). *MoralityGym: A Benchmark for Evaluating Hierarchical Moral Alignment in Sequential Decision-Making Agents*. DOI: 10.48550/arXiv.2602.13372. arXiv: 2602.13372 [cs.AI].

Spranca, Mark, Elisa Minsk, and Jonathan Baron (1991). “Omission and Commission in Judgment and Choice”. In: *Journal of Experimental Social Psychology* 27.1, pp. 76–105. DOI: 10.1016/0022-1031(91)90011-T.

Thomson, Judith Jarvis (1976). “Killing, Letting Die, and the Trolley Problem”. In: *The Monist* 59.2, pp. 204–217. DOI: 10.5840/monist197659224.

Wedekind, Claus and Manfred Milinski (2000). “Cooperation through Image Scoring in Humans”. In: *Science* 288.5467, pp. 850–852. DOI: 10.1126/science.288.5467.850.