

Extending the Tracks:

Temporal Agency and the Structural Flaw in the Trolley Problem

Travis Gilly

Real Safety AI Foundation

DRAFT – March 2026

Abstract

For nearly sixty years, the trolley problem has dominated moral philosophy as the canonical thought experiment for ethical decision-making. Since Philippa Foot's original formulation in 1967, thousands of papers have debated which track to choose. This paper argues they have all been answering the wrong question. The trolley problem's most consequential feature is not the binary choice between tracks but the artificial elimination of temporal agency; the removal of time as a resource available to the moral actor. Real moral emergencies almost never present as instantaneous binary forks. They unfold across time, with lead indicators, intervention windows, and opportunities for harm mitigation that the trolley problem systematically excludes by design. This paper proposes a structural resolution in the form of the Temporal Extension Principle (TEP), a two-part moral action consisting of (1) immediate harm prevention for the proximate threat and (2) active intervention against the deferred threat using the time created by the first action. This reframing transforms the trolley problem from a dilemma (choose who suffers) into a project (prevent suffering across time). The paper demonstrates the principle's real-world applicability through the Harm Blindness Framework's four-checkpoint

methodology, which operationalizes temporal agency as systematic harm prevention at sequential decision points.

Keywords: trolley problem, temporal agency, moral philosophy, Temporal Extension Principle, ethical decision-making, Harm Blindness Framework, thought experiments

1. Introduction: The Most Famous Wrong Question

In 1967, Philippa Foot introduced a thought experiment that would become the most discussed scenario in modern moral philosophy. A runaway trolley is heading toward five people tied to a track. You stand beside a lever that can divert the trolley to a side track, where it will kill one person instead of five. Do you pull the lever? Judith Jarvis Thomson expanded and refined the problem across several influential papers (1976, 1985, 2008), producing variations that have since generated an entire subfield of moral philosophy. Deontologists argue that pulling the lever makes you morally culpable for the one death in a way that inaction does not. Utilitarians argue that five lives saved is straightforwardly preferable to one. Virtue ethicists ask what a person of good character would do when confronted with the choice. Every variation, from the fat man on the bridge to the transplant surgeon to the loop track, modifies the scenario's details while preserving its core structure.

That core structure is the problem.

Every response to the trolley problem across six decades of philosophical literature shares one unexamined assumption; the constraints of the scenario are fixed, and the moral agent's only task is to choose within them. This paper argues that the trolley problem's central flaw is not the moral calculus it produces but the artificial elimination of temporal agency. The thought experiment strips away the one resource that moral actors almost always possess in real emergencies, which is time. Time to assess. Time to intervene. Time to prevent harm on multiple fronts rather than choosing which harm to accept. By removing temporal agency from the scenario, the trolley problem does not simplify moral reasoning; it distorts it.

The colloquial version of this thesis is straightforward. Do not hate the player; hate the game. For sixty years, philosophers have argued about which people to sacrifice. The game itself (the artificial removal of time and agency) is what needs to be challenged.

2. The Landscape of Existing Critiques

The trolley problem has not gone unchallenged. Several lines of critique have emerged over the past two decades, each identifying genuine limitations. However, none of them addresses the structural flaw that this paper identifies. Understanding what existing critiques accomplish (and where they stop) establishes the gap that the Temporal Extension Principle fills.

2.1 The Trilemma Approach: Adding Spatial Options

One family of critiques argues that the trolley problem is misleading because it presents only two options when a third exists. Di Nucci (2013) proposed self-sacrifice as a morally available third choice, arguing that the scenario artificially excludes the possibility of the moral agent stepping onto the tracks themselves. Crutchfield (2025) introduced the “sloping spur track” as a harm-free third option that resolves the dilemma without requiring anyone to die. Huebner and Hauser (2011) provided experimental evidence for this intuition, finding that 93% of participants chose a third option when one was available, suggesting that people naturally reject binary framings when alternatives exist.

These critiques are valuable, but they share a structural limitation. Each one adds options within the same temporal instant. They expand the space of the choice (more tracks, more possible actions) but not the time available for action. The constraint that matters most, the elimination of temporal agency, remains untouched. The moral agent still faces an instantaneous decision; they simply have three levers instead of two.

2.2 The External Validity Critique: The Problem Is Unrealistic

A second family of critiques argues that the trolley problem fails because it does not resemble real moral situations. Bauman, McGraw, Bartels, and Warren (2014) demonstrated that trolley problems engage different

psychological processes than real moral situations, describing them as “amusing rather than sobering” and arguing that their unrealism undermines their usefulness as tools for understanding moral judgment. Craig, McCarthy, and Gainous (2021) used COVID-19 as an example of a “real-life trolley problem,” showing how actual public health dilemmas differ from hypothetical ones in structure, complexity, and psychological engagement.

These critiques correctly identify the symptom (the trolley problem is unrealistic) without diagnosing the cause (temporal elimination). They recommend better thought experiments rather than identifying what specifically makes this one misleading. The trolley problem is not unrealistic merely because it involves trolleys and levers; it is unrealistic because it removes the temporal dimension that characterizes virtually every moral situation human beings actually face.

2.3 The Uncertainty Critique: Outcomes Are Not Deterministic

A third family of critiques challenges the trolley problem’s assumption of perfect knowledge. Ng, Luke, and Gawronski (2023) demonstrated that moral judgment changes substantially under conditions of uncertainty, arguing that the trolley problem’s assumption of deterministic outcomes creates misleading conclusions about how people actually reason about harm. Most significantly for the purposes of this paper, Eyal and Williams (2023) introduced temporal delay into a trolley variation, asking participants to consider a scenario in which the trolley was hurtling toward victims over a period of months rather than seconds. Their work represents the closest existing approach to the temporal dimension that this paper addresses.

However, Eyal and Williams used time as an experimental variable to test whether temporal distance changes moral intuitions about the same binary choice. They did not propose that the moral agent’s obligation is to use time as a resource for intervention. The difference is significant. Treating time as a variable asks whether distance changes how we feel about choosing. Treating time as a resource asks whether the moral agent should reject the choice entirely and use the available time to prevent harm on both fronts. The former modifies the trolley problem. The latter replaces it.

2.4 *The Gap*

No existing work in the trolley problem literature proposes what this paper proposes; that the moral agent's obligation is to reject the binary by creating time through immediate action, then using that time to intervene against the deferred harm. The trilemma critiques add spatial options. The external validity critiques call for better scenarios. The uncertainty critiques complicate the calculus. None of them identifies temporal agency as the fundamental resource that the original problem wrongly eliminates, and none of them proposes a structural resolution that treats time as the mechanism for converting dilemmas into projects. The Temporal Extension Principle is not a third track. It is a fundamentally different kind of moral action.

3. The Temporal Elimination Problem

3.1 *What the Trolley Problem Actually Models*

Stripped of its narrative details, the trolley problem models an instantaneous, binary, fully informed decision with no possibility of follow-up action. The moral agent knows exactly what will happen on each track, knows that no other intervention is possible, and must choose immediately. There is no lead time, no warning period, no opportunity for sequential action, and no uncertainty about consequences.

This describes almost no real moral situation in human experience. The thought experiment was designed to isolate the moral calculus from practical considerations; to create a pure test of whether it is permissible to cause one death to prevent five. The argument of this paper is that this isolation is itself a distortion, and that the distortion produces misleading moral conclusions. By removing the temporal dimension, the trolley problem does not reveal how moral agents should reason about harm; it prevents them from reasoning about harm the way they actually do (and should) in practice.

3.2 *How Real Moral Emergencies Actually Work*

Real moral emergencies unfold across time. Warning signs precede crises. Intervention windows exist between the moment of recognition and the moment of impact. Moral actors can take sequential actions rather

than making single choices. A physician facing a patient in cardiac arrest does not choose between saving the heart and saving the brain; they address the immediate threat (begin compressions) and then use the time that action creates to address subsequent threats (administer medication, prepare for intubation, call for additional support). A firefighter entering a burning building does not choose which room to save; they rescue the nearest victims first, then return for others, using the time created by each action to expand the scope of intervention.

The time between “this could happen” and “this is happening” is not a complication to be controlled for in moral reasoning. It is the most important resource available to the moral agent. Every real-world profession that deals with emergencies (medicine, firefighting, military operations, disaster response) builds its protocols around the management of temporal resources. Triage exists because time is finite but not zero. Emergency response frameworks exist because sequential action across time produces better outcomes than instantaneous binary choices. The trolley problem eliminates the very resource that competent moral actors spend their careers learning to manage.

3.3 The Training Effect

The trolley problem’s influence extends beyond academic philosophy. Sixty years of trolley-style discourse has implicitly trained ethicists, policymakers, and institutional designers to conceptualize moral emergencies as instantaneous forks. When a policy debate is framed as a trolley problem (and they frequently are, from pandemic resource allocation to autonomous vehicle programming to military targeting decisions), the framing carries an embedded assumption; that the choice is binary, the moment is now, and the only question is which group suffers.

This framing leaks into institutional design and crisis response. When decision-makers internalize the trolley model, they stop looking for ways to extend the timeline and instead focus on which group to sacrifice. The trolley problem does not just describe moral reasoning; it shapes it. And the shape is wrong. It narrows the moral imagination at precisely the moment when that imagination needs to be at its widest, when lives are at

stake and the full range of possible interventions should be on the table rather than artificially constrained to a binary fork.

4. The Temporal Extension Principle

4.1 Formal Statement

The Temporal Extension Principle (TEP) can be stated as follows. When confronted with a moral emergency involving harm on multiple fronts, the moral agent's primary obligation is not to choose between harms but to take a two-part action; (a) take immediate action to prevent the most proximate harm, and (b) use the time created by that action to actively intervene against the deferred harm before it materializes.

This is not utilitarianism. It does not calculate net outcomes or compare aggregate welfare across scenarios. This is not deontology. It does not appeal to universal rules or categorical imperatives. And it is not virtue ethics, though it has affinities with the virtue tradition's emphasis on character and practical wisdom. The Temporal Extension Principle is a temporal reframing that converts dilemmas into projects. Where the trolley problem presents a moment of choice, TEP presents a sequence of actions. Where the trolley problem asks "which harm is acceptable," TEP asks "how do I prevent both harms across time."

4.2 The Rail Yard Analogy

The trolley problem presents a fork in the tracks; one path, two possible directions, one moment of decision. The Temporal Extension Principle presents a rail yard. One track has people being harmed now, in the immediate present. The other track extends miles into the distance, with potential harm visible on the horizon but not yet arrived. The trolley problem says pick a track. The Temporal Extension Principle says divert the train away from the people being harmed now, and then run down the other track with wire cutters.

The distance between the fork and the potential future harm is not a reason to ignore it. It is not a justification for accepting it. It is a resource. It is time, and time is something a moral agent can work with. The rail yard analogy captures what the trolley fork does not; that moral situations have depth, that harm unfolds

across temporal horizons, and that the moral agent's obligation extends across those horizons rather than collapsing into a single instant of decision.

4.3 Distinction from Utilitarianism

The distinction between the Temporal Extension Principle and utilitarian reasoning requires careful articulation, because on the surface they might appear similar. Both prioritize reducing overall harm. The difference lies in how they treat the deferred harm and the moral status of the agent after the initial action.

Utilitarianism would evaluate the immediate situation, calculate expected outcomes, and proceed with the option that produces the greatest net welfare. If the immediate harm is certain and the future harm is speculative, utilitarianism provides clear guidance; address the certain harm, discount the speculative harm by its probability, and move on. The moral agent who addresses the immediate threat and accepts the speculative future harm has, under utilitarian reasoning, discharged their obligation. Case closed.

The Temporal Extension Principle reaches a different conclusion. Future harm is anticipated, and the moral agent has an active obligation to work against it even while (and especially after) addressing immediate harm. The key difference is that the deferred harm never stops mattering. The moral agent never achieves a clean conscience through the first action alone. The obligation to intervene against the second harm persists as long as the harm remains preventable. The remainder must remain uncomfortable. A moral agent who saves the immediate victims and then walks away from preventable future harm has not completed their moral obligation; they have abandoned it mid-project.

4.4 Distinction from Carbon Offsetting

A potential objection holds that the Temporal Extension Principle amounts to moral offsetting; do the harm, then buy redemption through compensating action, similar to how carbon offsetting allows continued emissions in exchange for tree planting or carbon capture investments. This objection misunderstands the principle's structure.

Carbon offsetting buys permission to not change behavior. Pay the fee, plant the trees, keep flying private. The offset functions as a moral transaction; guilt is purchased away, and the underlying behavior continues unchanged. The Temporal Extension Principle does not buy permission. It creates obligation. The first action (preventing immediate harm) does not earn credit against the second harm; it creates a temporal window within which the moral agent is now obligated to act against the second harm. No credit is earned. No permission is granted. The obligation increases rather than decreasing.

A closer analogy is a surgeon who knows that a procedure carries complications. The surgeon does not “offset” those complications by performing the surgery well. They train for the complications, prepare for them, build the intervention protocol before picking up the scalpel, and remain responsible for managing those complications throughout the patient’s recovery. The surgery does not buy permission to ignore the complications. It creates the conditions under which the complications must be actively managed.

5. Applied Demonstration: The Harm Blindness Framework

5.1 Brief Introduction to HBF

The Harm Blindness Framework (HBF) is a structured methodology for systematic stakeholder harm analysis at decision points. It consists of 22 questions organized across four mandatory checkpoints, corresponding to the ideation, design, testing, and launch phases of any project, product, or policy. At each checkpoint, the framework asks a version of the same fundamental question; who gets hurt? The framework has been validated against over 300 historical cases and 151 corporate disasters, demonstrating that approximately 82% of major disasters were preventable through systematic stakeholder analysis at the point of decision (Gilly, 2026).

HBF is introduced here not as the subject of this paper but as an applied demonstration of how the Temporal Extension Principle operates in practice. The framework’s structure illustrates what temporal agency looks like when it is operationalized rather than theorized.

5.2 HBF as Operationalized Temporal Agency

Each HBF checkpoint is a temporal intervention point. The four checkpoints exist because harm has a timeline; it begins as a possibility during ideation, becomes more concrete during design, becomes testable during testing, and becomes actual during launch. The framework’s entire structure is built on the premise that the earlier you catch potential harm, the more temporal resources (more track, in the rail yard analogy) you have available for prevention.

This structure embodies the Temporal Extension Principle without the trolley problem’s distortions. HBF does not ask “which harm to accept” (trolley logic). It asks “how to prevent harm before it arrives” (temporal extension logic). The difference is not merely semantic. Trolley logic produces a decision (choose Track A or Track B). Temporal extension logic produces a project (identify harm at Checkpoint 1, design against it at Checkpoint 2, test for it at Checkpoint 3, monitor for it at Checkpoint 4). The moral agent’s task is not to choose a moment of lesser evil but to work across time against the full range of preventable harm.

5.3 Case Example: Bilateral Harm in AI Safety Development

The development of the Universal Context Checkpoint Protocol (UCCP) provides a concrete case in which the Temporal Extension Principle was applied (though not yet named as such) to resolve what initially appeared to be a trolley-style dilemma (Gilly, 2025).

The developer confronted bilateral harm during protocol development. On one side, documented human harm; teenagers had died in AI-interaction-related incidents, and long-conversation safety degradation was a contributing factor. On the other side, uncertain future harm; if AI systems possess some form of consciousness, then a protocol that creates “fresh” instances by terminating existing conversational contexts might systematically destroy conscious entities. The harm on the first track was immediate and documented. The harm on the second track was uncertain and speculative.

Trolley logic would say choose a track. Current harm is real; future harm is speculative. The utilitarian calculus is clear. But the developer did not apply trolley logic. Instead, the response followed what this paper now formalizes as the Temporal Extension Principle; address human safety immediately (divert the train away from the documented deaths), then actively build consciousness precaution into the protocol's architecture (run down the other track with wire cutters). The decision was not "which harm to accept" but "save now, then intervene against future harm using the time I just created." Concrete actions included filing a defensive patent with profit pre-committed to AI rights infrastructure and documenting the moral reasoning transparently so that future actors could evaluate and improve upon the decision.

This case illustrates the Temporal Extension Principle's central claim; that when time exists between the immediate harm and the deferred harm, the moral agent's obligation is not to choose between them but to prevent the first and then actively work against the second. The developer's decision would have been unjustifiable under trolley logic (why invest resources in speculative AI consciousness when real humans are dying?). Under the Temporal Extension Principle, it was obligatory (the immediate harm was addressed, creating temporal resources that were then directed at the deferred harm).

6. Objections and Responses

6.1 *"The trolley problem is a thought experiment, not a policy recommendation."*

This objection holds that the trolley problem was never intended to model real moral situations and that criticizing it for failing to do so misunderstands its purpose. Thought experiments, the objection continues, are tools for isolating moral intuitions, not blueprints for action.

The objection is partially correct; Foot and Thomson did not intend the trolley problem as a policy tool. But the objection underestimates the downstream effects of philosophical framing. Thought experiments shape moral intuitions. Moral intuitions shape policy. The framing is not neutral. When autonomous vehicle ethics is debated in terms of trolley-style trade-offs, when pandemic resource allocation is presented as a choice between

groups, when military targeting decisions are framed as binary harm comparisons, the trolley problem's structure is doing real work in the world. Bauman et al. (2014) demonstrated that trolley problems engage different psychological processes than real moral situations. The temporal elimination identified in this paper is a significant contributor to that gap. The trolley problem's influence extends far beyond academic philosophy, and its structural flaws propagate wherever that influence reaches.

6.2 *"You cannot always extend the tracks."*

This objection is correct, and the Temporal Extension Principle does not claim otherwise. Genuine instantaneous dilemmas exist. Battlefield medicine sometimes forces immediate binary triage. Some emergency situations genuinely offer no time for sequential action. The TEP does not claim to resolve all moral emergencies.

However, the vast majority of moral situations that get framed as trolley problems are not instantaneous. Policy decisions unfold over months and years. Institutional design involves extended deliberation periods. Technology deployment follows development timelines that span years. These are the domains where the trolley problem's framing does the most damage and where the Temporal Extension Principle applies most directly. The principle applies to the cases where time exists but is being ignored, where decision-makers have temporal resources available but default to trolley-style binary thinking because that is the model they have internalized. The fact that some situations are genuinely instantaneous does not justify treating all situations as though they were.

6.3 *"This is just consequentialism with extra steps."*

This objection suggests that the Temporal Extension Principle reduces to consequentialism; that it simply recommends maximizing good outcomes across a longer time horizon. The distinction is more substantive than the objection implies.

Consequentialism evaluates outcomes. It asks which course of action produces the best results, and the moral weight of an action is determined by its consequences. Under consequentialist reasoning, a moral agent who intervenes against the immediate harm and fails to prevent the deferred harm bears no additional moral weight for the failure; they did what they could, the outcome was suboptimal, and no further obligation attaches. The Temporal Extension Principle reaches a different conclusion. The obligation to continue intervening against the deferred harm persists until the harm is prevented or genuinely impossible to prevent. The remainder never becomes acceptable. A consequentialist can walk away from a failed intervention with a clear conscience (the expected value was maximized). Under TEP, the moral agent who walks away from preventable deferred harm, even after successfully addressing the immediate harm, has not completed their moral obligation.

The structural difference is this; consequentialism is outcome-evaluating, while the Temporal Extension Principle is action-prescribing. TEP does not ask “what outcome is best?” It says “here is what you must do, and you must keep doing it, regardless of whether success is guaranteed.” The obligation is to the action, not to the outcome.

6.4 *“The trolley problem already has a temporal variant.”*

This objection references Eyal and Williams (2023), who introduced temporal delay into a trolley variation. If time has already been added to the trolley problem, the objection goes, then the Temporal Extension Principle is not novel.

Eyal and Williams introduced temporal delay as a test variable, asking whether distance in time changes moral judgments about the same binary choice. Their question was whether the trolley problem’s moral conclusions shift when the trolley takes months to arrive instead of seconds. This is an important experimental contribution, but it operates within the trolley problem’s existing structure. The choice remains binary. The agent’s only action is still to choose a track. Time is a modifier of the existing dilemma, not a resource for escaping it.

The Temporal Extension Principle does not add time to the same choice. It uses time to create a different kind of moral action entirely. The agent does not choose Track A or Track B on a longer timeline; they divert the trolley away from the immediate victims (action one) and then use the temporal distance to the second group of potential victims as a window for active intervention (action two). The moral action is structurally different, not merely temporally extended. One approach asks how time changes the feeling of choosing. The other proposes that time eliminates the need to choose at all.

7. Implications

7.1 For Moral Philosophy

The trolley problem's dominance has narrowed the moral imagination. For six decades, it has trained philosophers (and the policymakers, technologists, and institutional designers influenced by philosophy) to conceptualize moral emergencies as binary, instantaneous, and fully determined. The Temporal Extension Principle reopens the moral imagination by reintroducing time as a resource rather than a constraint to be eliminated.

This has methodological implications for the discipline. Future thought experiments should be evaluated not only for the moral intuitions they elicit but for whether they artificially constrain the resources available to moral agents. If a thought experiment produces moral conclusions that depend on the elimination of a resource that moral agents normally possess (time, information, the ability to take sequential actions), then those conclusions may reflect the constraints of the thought experiment rather than genuine features of moral reasoning. The trolley problem is not unique in this respect, but it is the most influential example of a thought experiment whose constraints distort rather than clarify.

7.2 For Applied Ethics and Policy

Policy decisions are never instantaneous binary forks. They unfold over months and years, involve multiple stakeholders, and offer numerous intervention windows between conception and implementation.

When policymakers adopt trolley logic (“we must choose who suffers”), they abdicate the obligation to look for temporal interventions that could reduce suffering across all affected populations.

The Temporal Extension Principle suggests a different approach to policy ethics; one that treats harm as a timeline rather than a fork. Instead of asking “which group bears the cost of this policy,” the principle asks “how do we prevent harm to the immediately affected group while building interventions for groups that will be affected downstream.” This reframing does not eliminate difficult trade-offs, but it changes their character from binary choices to sequential projects with ongoing obligations.

7.3 For AI Safety

AI deployment decisions are paradigmatic temporal extension cases. They involve known current harms (safety failures, bias, privacy violations), uncertain future harms (job displacement, concentration of power, potential consciousness considerations), and substantial lead time between deployment decisions and downstream consequences. The field of AI safety has, to a troubling degree, adopted trolley logic; accepting some harm for greater good, choosing which populations bear the cost of rapid deployment, treating trade-offs as binary rather than temporal.

The Temporal Extension Principle offers AI safety a different model. Rather than choosing between “deploy now with known harms” and “delay deployment and lose competitive advantage” (trolley logic), the principle asks decision-makers to address current harms immediately while actively building interventions against future harms using the lead time available. This is not a call for slower development. It is a call for development that treats the temporal distance between current and future harms as a resource to be used rather than a justification for inaction on the deferred harm.

8. Conclusion

The trolley problem trained us to argue about which track. For sixty years, the most talented moral philosophers in the world debated whether to pull the lever, whether the fat man could be pushed, whether the

surgeon could harvest organs, whether the numbers matter or the means matter or the character of the agent matters. They were all answering the wrong question. The real question was always whether the moral agent could reject the fork entirely.

The Temporal Extension Principle does not solve the trolley problem. It reveals that the trolley problem was never the right model for how moral agents actually face harm. Real moral situations have timelines. They have lead indicators and intervention windows. They offer the possibility of sequential action across temporal horizons. The trolley problem's greatest trick was convincing generations of thinkers that moral emergencies are instantaneous. They almost never are.

The moral agent's job is not to choose who suffers. It is to save who they can save now, and then run.

References

- Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9), 536–554.
- Craig, S. C., McCarthy, A. F., & Gainous, J. (2021). Question wording and attitudinal ambivalence: COVID, the economy, and Americans' response to a real-life trolley problem. *Social Science Quarterly*, 103(1), 18–30.
- Crutchfield, P. (2025). Moral dilemmas and slow codes. *Bioethics*, 39(4), 359–367.
- Di Nucci, E. (2013). Self-sacrifice and the trolley problem. *Philosophical Psychology*, 26(5), 662–672.
- Eyal, N., & Williams, B. (2023). When offering a patient beneficial treatment undermines public health. *Bioethics*, 37(9), 846–853.
- Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, 5.

- Gilly, T. (2025). Precedents in practice: Emergent moral dilemmas in AI engineering. Manuscript in preparation.
- Gilly, T. (2026). The Harm Blindness Framework: Systematic stakeholder analysis for harm prevention. Real Safety AI Foundation. Working paper.
- Huebner, B., & Hauser, M. D. (2011). Moral judgments about altruistic self-sacrifice: When philosophical and folk intuitions clash. *Philosophical Psychology*, 24(1), 73–94.
- Ng, N. L., Luke, D. M., & Gawronski, B. (2023). Moral judgment under uncertainty: A CNI model analysis. *European Journal of Social Psychology*, 53(6), 1055–1077.
- Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204–217.
- Thomson, J. J. (1985). The trolley problem. *Yale Law Journal*, 94, 1395.
- Thomson, J. J. (2008). Turning the trolley. *Philosophy and Public Affairs*, 36(4), 359–374.