

The Moral Grammar Exit

Eliminativist Moral Non-Realism, Pro-Social Evaluation,
and the Replacement of Moralized Control

Phil Stilwell

Independent Scholar

ORCID: 0009-0009-3725-8682

Revised August 2026

Central claim

Moralized verdicts often compress facts, attitudes, coordination demands, power relations, and recommendations into language that appears authoritative without showing its work. Compared with retaining that grammar under realism, expressivism, or fiction, explicit translation makes public disagreement more testable and accountable while preserving humane opposition and collective action. The case supports default replacement where those functions survive and the gain is real; it does not disprove moral realism or require a ban on ordinary emotional shorthand.

Preview: The replacement test

Moralized verdict	Translation burden	Public comparison	Bounded conclusion
"This is wrong" presents one charged verdict.	State the harm, interests, consent, trust, power, evidence, and proposed response.	Ask whether the explicit version preserves protection and coordination while improving scrutiny.	Retire the moralized form when translation keeps the useful work and reduces opaque authority; otherwise retain or revise it.

This is a defeasible policy for public reasoning, not a claim that every moral sentence has one hidden meaning or that nonmoral language is automatically neutral.

Visual preview:

Beyond the Moral Grammar Exit

Moralized Compression



Translation



Replacement Vocabulary



MORALIZED COMPRESSION

Vague terms like "wrong" or "evil" bundle and hide distinct social facts.

TRANSLATION

Decompressing moralized assertions into verifiable claims about human interaction.



REPLACEMENT VOCABULARY

No Moral Furniture

Evaluate harm, trust, preference, and power.

Read the image from left to right: a moralized verdict appears as a tangled bundle, translation separates its claims, and the replacement vocabulary names harm, preference, trust, and power directly. Those labels are prompts, not a complete theory: evidence, consent, alternatives, institutional effects, and a recommendation must still be supplied. The image gives disagreement visible handles; it does not show that every translation will persuade or that moral realism is false.

Abstract

Moral non-realism faces a practical question after its negative case is made: what should people say and do if moral predicates do not report a special realm of categorical facts? Moralized verdicts commonly combine several distinguishable acts—describing harm, expressing attitudes, invoking trust or coordination, positioning power, and recommending a response—while presenting the bundle as one authoritative property. This paper compares four policies toward that grammar: robust realist retention, expressivist retention, error-theoretic fictionalism, and explicit replacement. It develops a translation protocol that asks speakers to state affected interests, consent and exit, cooperative expectations, power relations, evidence, goals, and proposed interventions. The main result is conditional but substantive. When translation preserves warning, protection, coordination, and motivation while making assumptions and tradeoffs easier to inspect, public reasoning has a defeasible reason to retire the moralized form. Research on moral-emotional communication supplies a limited supporting strand: such language can spread rapidly within ideological networks and, in specific protest settings, track support for violence. Those findings do not show that all moral language is harmful or that replacement will always work. Strong objections also remain: realist authority may be genuine, expressivist grammar may efficiently coordinate attitudes, fiction may preserve useful commitments, and replacement vocabularies can smuggle values or burden vulnerable speakers. The appropriate conclusion is therefore a deployment rule, not universal linguistic prohibition: translate where the practical functions survive and accountability improves; retain or redesign the form where they do not. Eliminativism can preserve humane seriousness, but it must earn that claim in use.

Keywords: moral non-realism; moral error theory; moral abolitionism; expressivism; moral fictionalism; conceptual ethics; public reasoning; moralized language

1 The Question After Moral Error

Moral error theory says, roughly, that ordinary moral thought aims to represent objective prescriptive facts and systematically fails because no such facts exist. Mackie’s classic argument emphasizes the unusual authority and metaphysical character that moral properties would need to possess (Mackie 1977). Contemporary defenses refine the view and dispute how far the error extends (Olson 2014; Streumer 2017). Yet even a successful error theory would leave a second question unanswered: what policy should govern the language, institutions, and habits built around the alleged facts?

That second question is the focus here. The target is not every use of words such as *good*, *bad*, *right*, or *wrong*. People use them to joke, praise, warn, affiliate, teach, grieve, and summarize. The target is narrower: a *moralized public verdict* that presents an action, person, rule, or institution as categorically authorized or condemned while leaving its descriptive, affective, strategic, and political components compressed. The characteristic example is not merely “I hate this” but “This is wrong” used as though the predicate itself completes the public reason.

The primary comparator is explicit translation. Instead of treating wrongness as an additional feature hovering over injury, coercion, betrayal, and risk, the speaker states the relevant human and institutional facts, the interests at stake, the desired ends, and the proposed response. The comparison is not between morality and silence. It is between two ways of conducting evaluation: one through inherited moralized shorthand and one through decomposed claims that expose their evidence and aims.

The governing thesis is conditional. If translation preserves the legitimate practical work of the moralized verdict and produces a net gain in public accountability, then there is a defeasible reason to retire the moralized form in that context. The conclusion is stronger than a request for occasional clarification: it places the burden on the compressed verdict once an adequate explicit substitute exists. But it is weaker than a universal ban. Translation may fail, cost too much, or discard a function that participants reasonably need. Nor does the practical conclusion prove error theory true. A robust

realist could reject the starting diagnosis and still learn from the demand to show which public work the moral predicate adds.

This distinction matters because post-error policies diverge. Garner defends moral abolition rather than continuing a discourse one regards as false, while Joyce and Nolan, Restall, and West explore fictionalist strategies for retaining useful moral practice without ordinary belief (R. T. Garner 1994; R. Garner 2007; Joyce 2001; Nolan et al. 2005). Marks likewise proposes a language of desire as an alternative to morality (Marks 2013). The dispute is therefore not simply realism versus nihilism. It is a practical contest among rival linguistic policies.

The replacement principle

A moralized verdict should not receive extra public authority merely because it is phrased morally. Where its useful functions can be preserved by an explicit translation that makes harms, interests, evidence, power, and recommendations easier to inspect, public reasoners have a defeasible reason to use the translation instead.

2 What Moralized Compression Does

The claim that moral predicates compress discourse is not yet a theory of their complete meaning. Expressivists have long argued that moral judgments express attitudes, accept norms, and guide action rather than merely report special properties. Gibbard's norm-expressivism and Blackburn's quasi-realist program show how rich moral thought may be constructed from practical attitudes without reducing every utterance to a crude cheer or boo (Gibbard 1990; Blackburn 1998). Their sophistication is important: an eliminativist should not pretend that ordinary moral grammar contains only metaphysical error.

Compression instead names a public effect. Consider "The employer's surveillance is wrong." The sentence may carry at least six claims:

1. workers lose privacy and face predictable psychological or material harms;
2. the speaker and affected workers strongly oppose the practice;
3. consent is absent or structurally weak because exit is costly;
4. trust and cooperative expectations are damaged;
5. managers possess asymmetric power to collect and use information; and
6. the practice should be prohibited, limited, disclosed, or otherwise changed.

These claims can support one another, but they are not identical. Evidence might confirm the surveillance while leaving its psychological effects uncertain. Workers might accept some monitoring under negotiated conditions. A ban, audit rule, data-minimization requirement, or compensation scheme might serve different goals. Calling the practice wrong can be a useful alarm; it can also conceal which of these propositions is doing the work.

Compression becomes objectionable when the moral form is treated as self-authenticating. The speaker moves from a charged label to a demand without exposing the bridge. Disagreement then appears to concern one property even when participants contest different facts, weight different interests, or negotiate which concept should govern the case. Work in conceptual ethics asks which concepts we should use, while analyses of metalinguistic negotiation show that apparently first-order disputes may partly concern how a term ought to be deployed (Burgess and Plunkett 2013; Plunkett and Sundell

2013). These frameworks do not entail abolition, but they make vocabulary choice a legitimate object of evaluation.

Realists can resist the diagnosis. Enoch argues that robust normative truths are indispensable to deliberation, and Cuneo contends that attacks on moral normativity threaten an interconnected web that includes epistemic norms (Enoch 2011; Cuneo 2007). On such views, explicit facts about harm and preference do not exhaust the reason. A person can have decisive reason not to dominate another even when domination advances every desire the first person happens to possess. If that realist claim is true, replacement risks deleting rather than clarifying authority.

The reply should not be a slogan. The practical argument can proceed conditionally: even if participants dispute whether categorical reasons exist, public deliberation improves when they identify the empirical and institutional grounds that a policy must address. The realist remains free to add that these grounds instantiate genuine moral reasons. What the realist may not do is treat the predicate alone as a substitute for explaining which reasons apply and why. Translation therefore imposes a burden of articulation without assuming at the outset that realism is false.

3 A Translation Protocol

An adequate replacement cannot consist of deleting *wrong* from a sentence. It must preserve the functions that made the sentence worth uttering. Table 1 gives a practical protocol. The categories are prompts rather than metaphysical atoms; a single claim may appear in more than one row.

Table 1: What an explicit translation asks a speaker to supply.

Component	Public question	Inspect-able content	What it prevents
Harm and benefit	Who experiences what effects, with what severity, duration, probability, and distribution?	Injury, deprivation, trauma, opportunity, risk, repair, and uncertainty.	Letting a charged label replace causal evidence.
Attitudes and interests	Who wants, fears, rejects, or prioritizes what, and how strongly?	Preferences, commitments, identities, projects, and conflicts among them.	Presenting one group's concern as already shared by everyone.
Consent and exit	What was agreed, under which options, and at what cost of refusal or departure?	Information, bargaining position, dependency, revocability, and feasible alternatives.	Treating formal permission as sufficient when power makes refusal unrealistic.
Trust and coordination	Which expectations, promises, or institutions are sustained or damaged?	Reliability, reciprocity, information flow, stability, and generalized effects.	Ignoring the social conditions that make cooperation possible.
Power and enforcement	Who sets the rule, bears its cost, can contest it, and controls sanctions?	Domination, exclusion, vulnerability, capture, accountability, and remedy.	Hiding coercion inside allegedly self-authorizing norms.
Recommendation	Given stated ends and constraints, what response is proposed and why?	Policy, refusal, protection, sanction, redesign, compromise, and review condition.	Moving from disapproval to intervention without a visible bridge.

The protocol separates three layers that moral speech often merges. The *descriptive layer* concerns what happened and what is likely to follow. The *evaluative layer* concerns what participants care about and which tradeoffs they accept. The *action layer* concerns what should be done relative to those ends and constraints. Separation does not make the evaluative layer disappear; it stops that layer from borrowing the appearance of an observed property.

Take a forceful condemnation of torture. Translation does not weaken it to “I dislike torture.” It can state that torture deliberately inflicts extreme suffering, uses domination to destroy agency, produces unreliable information, traumatizes victims and agents, corrupts institutions, and creates fear under

which ordinary cooperation is impossible. It can add that those exposed to the practice demand security against it and that public institutions committed to bodily integrity and accountable power should prohibit, investigate, and remedy it. The translation is longer because the case contains more than one reason. Its force comes from the disclosed injuries, relationships, commitments, and institutional consequences, not from a mysterious extra property.

The conditional form is not an embarrassment. Engineering advice is conditional on safety, performance, cost, and user aims. Public-health recommendations are conditional on reducing illness under constraints of evidence and liberty. Acknowledging the condition lets participants contest it. If a government prizes domination, or a speaker values only in-group welfare, translation exposes that conflict instead of pretending that both sides already share one moral authority.

This protocol also disciplines recommendations. Facts about harm do not mechanically entail one policy. A ban may protect some people while driving a practice underground; punishment may deter or may reproduce domination; disclosure may help informed consent while imposing unequal costs. Explicit language forces the proposal, alternative, evidence, and review condition into view. Eliminativism earns its constructive name only when it improves this practical reasoning rather than merely changing labels.

4 Four Policies Toward Moral Grammar

The practical choice is easier to evaluate when the alternatives are stated in their strongest forms. Table 2 compares four discourse policies. None follows automatically from a single metaethical premise.

Table 2: Competing policies for moralized language.

Policy	What it preserves	Principal strength	Principal burden
Robust realist retention	Truth-appt moral judgments and categorical reasons.	Explains why some demands bind independently of desire or agreement.	Defend the ontology, epistemology, and action-guiding role of irreducible norms.
Expressivist retention	Moral grammar as expression, norm acceptance, and practical coordination.	Preserves familiar discourse without treating it as simple property reporting.	Explain why quasi-realist language does not recreate the opacity or authority inflation under challenge.
Error-theoretic fictionalism	Familiar moral practice used without ordinary belief in moral facts.	Retains motivational and coordinating benefits while accepting error theory.	Show that the fiction is stable, transparent, and preferable to explicit alternatives rather than deceptive or confusing.
Eliminativist replacement	Evaluation through harm, interests, consent, trust, power, evidence, and stated recommendations.	Makes assumptions and tradeoffs directly contestable.	Show that translation preserves protection, solidarity, motivation, and manageable communication costs.

4.1 Realist retention

Robust realism offers the strongest challenge because it denies that translation captures all relevant facts. If irreducible normative truths are real and epistemically accessible, then a complete account of torture or betrayal may include the truth that it is wrong. Enoch's deliberative-indispensability argument and Cuneo's parity between moral and epistemic normativity show why a non-realist cannot dismiss the view as mere supernaturalism (Enoch 2011; Cuneo 2007).

The replacement argument therefore does not claim that its practical success entails realism's falsity. It makes two narrower points. First, a moral predicate that tracks a real reason would still need application conditions. Public agents must explain why this case instantiates the reason and what response follows. Second, when every action-guiding consideration can be made explicit and the moral predicate changes no prediction, priority, or remedy, the predicate has no demonstrated additional public work. The realist may retain it as true; the public case should not depend on its unexplained authority.

4.2 Expressivist retention

Expressivists argue that moral language earns its place by expressing practical stances and coordinating norms. Gibbard treats normative judgment through acceptance of systems of norms, while Blackburn develops a quasi-realist practice rooted in attitudes but capable of familiar ethical argument (Gibbard 1990; Blackburn 1998). This is not a failed attempt at factual description. It is a rival account of what speakers are doing.

The eliminativist objection is pragmatic: the familiar surface grammar can invite users to treat attitudes and social demands as discovered categorical facts. That risk is not inevitable. An expressivist community could teach its semantics openly and require the same articulation demanded here. If it did so while retaining brevity and motivation, replacement might have no net advantage. The dispute must then be settled by comparative performance, not by declaring one vocabulary metaphysically pure.

4.3 Fictionalist retention

Fictionalism accepts something like error theory while preserving moral discourse for its useful effects. Joyce explores how moral thought might function as a useful fiction, and Nolan, Restall, and West defend fictionalism as a way to retain moral phenomena without objective prescriptivity (Joyce 2001; Nolan et al. 2005). The attraction is clear: communities need not rebuild ordinary speech simply because philosophers reject its literal commitments.

Garner's abolitionist response is that a discourse regarded as systematically false risks encouraging deception and epistemic confusion (R. Garner 2007). That criticism is strongest when the fiction is not publicly marked. A transparent game may not deceive, but then speakers need a shared account of when they are asserting, accepting, pretending, or merely coordinating. The practical question returns: is this machinery clearer and more reliable than direct translation? Fictionalism wins if it preserves indispensable functions at lower cost without encouraging literal reification. Replacement wins if the fiction repeatedly leaks back into authority claims.

4.4 Replacement

Replacement has the cleanest semantics and the largest empirical burden. Garner and Marks argue that life beyond moral judgment can preserve serious evaluation, but the claim cannot be secured by examples chosen by sympathetic authors (R. T. Garner 1994; Marks 2013). It requires evidence that people can coordinate, protect vulnerable parties, sustain commitments, and motivate costly action without treating moral predicates as categorical facts.

The proposal should therefore be understood as staged conceptual engineering. Burgess and Plunkett's central question—which concepts should we use?—invites comparison of representational and practical performance, not a one-time decree (Burgess and Plunkett 2013). Institutions can test replacement in settings where reasons must already be recorded: policy analysis, mediation, risk review, consent procedures, workplace rules, and public justification. The standard is not whether every participant likes the result. It is whether relevant claims become easier to assess without disabling the functions the moralized form served.

5 Amplification, Conflict, and the Limits of the Evidence

The source intuition that moral language amplifies emotion has empirical support in some settings, but it must be stated carefully. Brady and colleagues analyzed social-media discussions of polarizing issues and found that moral-emotional language was associated with greater diffusion, especially within ideological group boundaries (Brady et al. 2017). Mooijman and colleagues combined social-media data from the 2015 Baltimore protests with experimental work; they reported associations among moralization, perceived network convergence, support for violence, and protest violence (Mooijman et al. 2018).

These studies contribute one strand to the practical case. They show that moral-emotional framing can affect transmission and may interact with group alignment and escalation. They do not show that moral vocabulary alone caused every measured outcome, that nonmoral language would have prevented it, or that the same framing is harmful in every context. Moral-emotional communication can also recruit help, signal solidarity, and make neglected injuries visible. A language that spreads rapidly is not defective merely because it spreads.

The relevant inference is comparative. If moralized language increases diffusion by activating identity and emotion, that may be useful for mobilization but costly for cross-group inquiry. Translation can slow the move from alarm to condemnation by asking for injuries, evidence, power relations, and proposed remedies. Slower is not always better; emergencies require concise warnings. The practical question is when speed serves protection and when it bypasses interpretation.

This yields a testable prediction. In contexts of contested public policy, an explicit translation should improve agreement about factual subclaims, reveal where values or forecasts diverge, and reduce false consensus about the meaning of a verdict. It might nevertheless reduce motivation, increase message length, or provide strategic actors with new ways to obscure responsibility. Studies comparing moralized and translated formulations should therefore measure not only civility, but comprehension, recall, willingness to protect, perceived legitimacy, polarization, policy agreement, and resistance to manipulation.

The empirical evidence is presently insufficient for universal abolition. That limit strengthens rather than dissolves the argument. It converts a philosophical preference into a research program with possible failure conditions. If moralized forms reliably protect vulnerable parties, sustain cooperation, and improve decisions after their risks are controlled, replacement should retreat. If explicit forms preserve those gains while improving accuracy and contestability, the abolitionist case becomes stronger.

6 Five Stress Tests

6.1 Does replacement smuggle values?

Yes, if it pretends that harm, consent, trust, or non-domination select themselves. The protocol is not value-free. It asks speakers to state values rather than hide them. “Reduce severe involuntary suffering” is a substantive priority. “Preserve trust” may conflict with privacy or dissent. “Increase freedom” depends on whose options count and how power is measured. Translation succeeds by making those commitments visible, not by escaping evaluation.

This answer also blocks a common overclaim. Non-realism does not entail that every preference deserves equal institutional weight. Agents can oppose sadistic or exclusionary preferences because of the harms, dependencies, reciprocal risks, and social arrangements they produce, and because the agents themselves commit to protections that conflict with those preferences. What non-realism withholds is the claim that a categorical property settles the conflict independently of all practical commitments.

6.2 Does abolition weaken protection for vulnerable people?

It can. Moral language supplies compressed signals of standing, violation, and solidarity. Rights talk, for example, can mark some interests as protected against ordinary bargaining. Removing that language without institutional substitutes would be reckless. A successful translation must therefore preserve threshold protections: who may not be exposed to what risks, which interests are non-negotiable under a stated charter, who has standing to object, and what remedies follow.

The eliminativist response is not to replace “This violates her rights” with a private preference. It is to state the protected interest, the institutional commitment, the power asymmetry, the rule that blocks tradeoffs, and the remedy. If the longer account weakens compliance or makes abuse easier to excuse, the translation fails the preservation test. The burden belongs to the reformer.

6.3 Can people coordinate without moral shorthand?

Often yes, but not costlessly. Teams use technical standards, contracts, procedures, warnings, and shared aims without first resolving moral realism. Yet ordinary moral terms are compact and widely learned. A requirement to unpack every small judgment would paralyze conversation.

The proposal is therefore a default for consequential public justification, not a policing rule for every utterance. “That was wrong” may open a conversation. It should not end one when sanctions, policy, exclusion, or coercion are at stake. Translation effort should rise with the action demanded and with reasonable disagreement about the verdict’s content.

6.4 Does explicit language motivate strongly enough?

Perhaps not in every setting. Moral emotions can recruit costly action, and the social-media evidence shows their transmission power (Brady et al. 2017). But motivation bought through opacity may also misfire. A replacement culture would need narratives, commitments, loyalties, empathy, and institutions; a list of neutral facts is insufficient.

Eliminativism does not require affective flatness. Speakers may say, “I am furious because this policy traps families in avoidable danger, and I will help stop it.” The sentence expresses emotion, identifies a cause, and announces action. Its force is not weaker merely because it omits *evil*. The open empirical question is whether such formulations can sustain collective action across different contexts.

6.5 Will bad actors exploit translation?

They will try. A powerful institution can bury coercion under data, redefine harm, manufacture consent, or demand endless evidence from people already injured. Moralized protest may sometimes be the only signal strong enough to break that frame. Translation is not a guarantee of good faith.

The protocol therefore includes power, standing, burden of proof, and feasible exit. It also allows provisional action under uncertainty. “There is credible evidence of severe risk; the exposed group lacks meaningful consent; pause the practice while an independent review tests the claim” is explicit without waiting for certainty. Replacement should make control visible, not convert public reasoning into technocratic delay.

7 A Deployment Rule

The paper’s practical recommendation can now be stated as a six-part test. Before using a moralized verdict to justify consequential public action, ask:

1. **Unpackability:** Can the verdict’s action-guiding content be stated through evidence, effects, interests, consent, coordination, power, and a recommendation?

2. **Function preservation:** Does the translation preserve warning, protection, solidarity, motivation, and coordination that are legitimate in this context?
3. **Accountability gain:** Does it make disputed assumptions, tradeoffs, and enforcement choices easier to identify and challenge?
4. **Burden control:** Are the time, literacy, and emotional costs manageable, especially for vulnerable speakers?
5. **Manipulation resistance:** Does the form expose rather than hide who controls evidence, definitions, sanctions, and opportunities to exit?
6. **Reviewability:** Can participants specify what result would show that the translation policy is failing?

Table 3: The deployment decision is conditional rather than absolute.

Observed pattern	Recommended policy	Reason
Functions preserved; accountability improves; costs manageable	Prefer explicit replacement	The moralized form adds no demonstrated public benefit and retains opacity risk.
Functions preserved; no accountability gain	Permit either form	Reform would create switching costs without a shown advantage.
Accountability improves; a vital function is lost	Redesign and retest	Clarity does not compensate for disabled protection or coordination.
Translation burdens vulnerable speakers or empowers gatekeepers	Retain concise protest language while requiring institutions to unpack their response	The party with less power should not carry the entire explanatory burden.
Moralized language reliably improves protection and decisions after controls	Retain it provisionally	A practical abolitionist must accept evidence against replacement.

This rule makes abolition local, revisable, and evidence-sensitive. It also preserves the paper’s polemical point. Moralized grammar should not be privileged merely because it is traditional, emotionally forceful, or spoken with certainty. When a speaker demands punishment, exclusion, or coercion, “wrong” is the beginning of an explanation, not its completion.

The result is post-moral rather than anti-human. People continue to care, condemn, protect, refuse, organize, and repair. What changes is the grammar of public authority. Harm must be shown; interests must be named; power must be located; recommendations must state their goals and costs. The replacement may be less grand. That is partly its virtue.

8 Conclusion

Eliminativist moral non-realism is incomplete when it merely denies moral facts. It must compare policies for what follows. Robust realism claims that translation omits genuine categorical reasons. Expressivism retains moral grammar as a sophisticated vehicle for attitudes and norm acceptance. Fictionalism preserves familiar practice without ordinary belief. Replacement asks whether their legitimate work can be performed more openly through claims about harm, interests, consent, trust, power, evidence, and proposed action.

The case for replacement is strongest where three conditions converge: the moralized verdict is unpackable, the translation preserves its protective and coordinating functions, and the explicit form

produces a net accountability gain. Moral-emotional communication research gives a limited reason to take opacity and amplification seriously, but it does not prove universal abolition. Translation can smuggle values, burden vulnerable speakers, weaken motivation, and invite technocratic control. Those are failure conditions, not footnotes.

The practical conclusion nevertheless carries force. In consequential public reasoning, a moral predicate should not receive authority for free. If the speaker can preserve the useful work while showing the injuries, commitments, power relations, evidence, and intervention, the explicit account deserves priority. If it cannot preserve that work, the eliminativist must redesign the proposal rather than declaring victory. Humane seriousness survives the moral grammar exit only when the replacement demonstrates it.

A Formal Reconstruction of the Central Thesis

The central argument is a conditional policy inference. It does not deduce moral non-realism, and it does not claim that every moralized utterance should disappear.

A.1 Ordinary-language reconstruction

- Premise 1.** If an explicit translation preserves the legitimate practical functions of a moralized public utterance and yields a net gain in accountability, then there is a defeasible reason to retire the moralized form in that context.
- Premise 2.** For the class of utterances targeted in this paper—consequential public verdicts whose action-guiding content can be stated through harm, interests, consent, coordination, power, evidence, and recommendation—the translation can preserve those functions when the six-part deployment test is satisfied.
- Premise 3.** When that test is satisfied, the translation makes assumptions, tradeoffs, and enforcement choices more inspectable while keeping its burdens and losses manageable.
- Premise 4.** Therefore, for that class of utterances when the test is satisfied, there is a defeasible reason to retire the moralized form.

The conclusion is deliberately conditional. A realist may deny that translation captures all relevant reasons. An expressivist may show that retained moral grammar performs the functions better. An empirical study may find that replacement weakens protection or increases manipulation. Any of these would challenge a premise or show that the deployment test is not satisfied.

A.2 Symbols

- m : any moralized public utterance under consideration.
- m_0 : one particular utterance in the target class for which the deployment evidence has been assessed.
- $T(m)$: the explicit nonmoral translation of m .
- $P(m, T(m))$: the claim that $T(m)$ preserves the legitimate practical functions served by m .
- $G(m, T(m))$: the claim that $T(m)$ produces a net accountability gain over m after costs and risks are included.
- $R(m)$: the claim that there is a defeasible reason to retire m in favor of its translation.
- $\forall$: “for every”; $\wedge$ means “and”; $\rightarrow$ means “if-then”; and $\therefore$ marks the conclusion.

A.3 Symbolic form

The symbolic form is:

$$\begin{aligned} & \forall m [(P(m, T(m)) \wedge G(m, T(m))) \rightarrow R(m)] \\ & P(m_0, T(m_0)) \wedge G(m_0, T(m_0)) \\ & \therefore R(m_0). \end{aligned}$$

Plain-English annotation. The first line states the decision rule: whenever a translation both preserves the useful work and improves accountability overall, retirement is warranted by default. The rule does not say that every translation meets those conditions.

Plain-English annotation. The second line says that the particular case m_0 has passed both tests. This is the line that requires real conceptual and empirical work: protection, motivation, coordination, burden, and manipulation risk all matter.

Plain-English annotation. The final line applies the rule to that case. Because the reason is defeasible, new evidence about lost functions or hidden costs can overturn the practical recommendation without changing the logical form.

A.4 Validity, support, and exact rebuttal routes

The inference is deductively valid. The first line applies to every relevant utterance; the second says that m_0 satisfies its two conditions; the conclusion follows by applying the general rule and then using modus ponens. Validity means only that the conclusion follows if the premises are true.

The premises have different kinds of support. The first is a conceptual policy principle: it expresses why a form that preserves the benefits and improves accountability should normally be preferred. The second is empirical and context-sensitive. The translation protocol, strategy comparison, bounded communication evidence, and five stress tests support it, but they do not prove it for every institution or culture. A critic can rebut the conclusion in four precise ways: show that m_0 cannot be adequately unpacked; show that $T(m_0)$ loses a legitimate practical function; show that the claimed accountability gain disappears after costs or power effects are counted; or reject the decision rule because moralized authority has an independently justified value. None of those rebuttals is answered merely by pointing out that the conclusion is logically valid.

References

- Blackburn, Simon (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198247852.001.0001.
- Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel (2017). “Emotion Shapes the Diffusion of Moralized Content in Social Networks.” In: *Proceedings of the National Academy of Sciences* 114.28, pp. 7313–7318. DOI: 10.1073/pnas.1618923114.
- Burgess, Alexis and David Plunkett (2013). “Conceptual Ethics I.” In: *Philosophy Compass* 8.12, pp. 1091–1101. DOI: 10.1111/phc3.12086.
- Cuneo, Terence (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199218837.001.0001.
- Enoch, David (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199579969.001.0001.
- Garner, Richard (2007). “Abolishing Morality.” In: *Ethical Theory and Moral Practice* 10.5, pp. 499–513. DOI: 10.1007/s10677-007-9085-3.
- Garner, Richard T. (1994). *Beyond Morality*. Philadelphia: Temple University Press.
- Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.
- Joyce, Richard (2001). *The Myth of Morality*. Cambridge: Cambridge University Press. DOI: 10.1017/CB09780511487101.
- Mackie, John L. (1977). *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Books.
- Marks, Joel (2013). *Ethics without Morals: In Defence of Amoralism*. New York: Routledge. DOI: 10.4324/9780203093078.
- Mooijman, Marlon, Joe Hoover, Ying Lin, Heng Ji, and Morteza Dehghani (2018). “Moralization in Social Networks and the Emergence of Violence during Protests.” In: *Nature Human Behaviour* 2.6, pp. 389–396. DOI: 10.1038/s41562-018-0353-0.
- Nolan, Daniel, Greg Restall, and Caroline West (2005). “Moral Fictionalism versus the Rest.” In: *Australasian Journal of Philosophy* 83.3, pp. 307–330. DOI: 10.1080/00048400500191917.
- Olson, Jonas (2014). *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198701934.001.0001.
- Plunkett, David and Tim Sundell (2013). “Disagreement and the Semantics of Normative and Evaluative Terms.” In: *Philosophers’ Imprint* 13.23, pp. 1–37. URL: <http://hdl.handle.net/1885/27641>.
- Streumer, Bart (2017). *Unbelievable Errors: An Error Theory about All Normative Judgements*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198785897.001.0001.