

The Displacement Fallacy

Resolution Theory as an Interface Layer for Agency, Responsibility, and Normativity

H. R. Easton

Independent Researcher

Email: Hamiltoneaston92@gmail.com

1. Introduction: Why These Disputes Persist

In multiple domains, certain disagreements do not merely endure; they appear structurally resistant to resolution. These include free will vs determinism, neuroscience vs responsibility, moralism vs medicalisation in addiction, institutional harm with “no author,” and whether advanced AI should be treated as a responsible agent. The argumentative pattern is strikingly consistent. One side offers increasingly complete causal explanation of behaviour. The other insists that agency and responsibility remain real. When treated as rival explanations, the dispute becomes a zero-sum contest: either explanation must be incomplete for responsibility to survive, or responsibility must dissolve as explanation improves. Resolution Theory diagnoses this as a category mistake. Explanation and authorship are not rival explanatory quantities. Explanation answers how an outcome arose; authorship answers whether a resolution is attributable to an agent or institution. Allowing explanation to displace authorship is not merely conceptually confused but architecturally dangerous in modern systems, where responsibility is routinely diffused and laundered through procedure, incentives, and abstraction. This paper isolates the recurring inference behind these disputes and provides a minimal structural remedy.

2. The Displacement Fallacy

2.1 Definition

The Displacement Fallacy is the mistaken belief that: As the density of causal explanation (Inputs → Evaluation) increases, the space for agential attribution (Resolution) must decrease. The fallacy is intuitive because it relies on a misleading metaphor: explanation “fills” causal space, leaving “no room” for agency. But agency is not a leftover causal ingredient. Under RT, authorship is not a gap in causation; it is a closure verdict attaching at a particular structural event: deliberative resolution.

2.2 Why the fallacy is tempting

The fallacy persists for three reasons. First, modern explanation is powerful. Neuroscience, predictive processing, reinforcement models, trauma histories, socioeconomic constraint, and institutional incentive analysis generate dense accounts of why certain outcomes are likely. Second, moral impulses toward compassion encourage explanation-as-exculpation: if we can explain a harmful action, it feels cruel to hold anyone responsible. Third—and most importantly—many theoretical frameworks treat responsibility as if it must compete with causation. This forces a false dilemma: preserve responsibility by inserting metaphysical gaps, or preserve causal completeness by dissolving responsibility. RT rejects the shared premise. Explanation does not displace authorship; it refines blame and mitigation while leaving responsibility structurally intact except in bypass cases.

3. Resolution Theory as an Interface Layer

3.1 The action pipeline

RT models action as: Inputs → Evaluation → Resolution → Action value, cost). (policy/intention). Causal explanation operates over inputs and evaluation. Attribution attaches at resolution. The framework explicitly diagnoses persistent disagreement as the repeated confusion of these explanatory roles.

- Inputs: biological dispositions, histories, incentives, pressures, constraints.
- Evaluation: these inputs showing up as reasons-for-me (salience, urgency,
- Resolution: deliberative closure into a committed course of action
- Action: execution downstream of closure.

3.2 Interface claim (verdict-types rather than rivals)

The central interface claim can be stated plainly: regarding the same event. A resolution event can be fully explainable and still attributable. Explanation can become arbitrarily detailed without ever summing into authorship, because authorship is not a causal remainder; it is a closure attribution.

- Causal explanation describes the history of the closure.
- Agential attribution identifies the locus of the closure.
- They are not rival accounts of the same thing, but different verdict-types

3.3 Why explanation cannot displace authorship

The core RT thesis is that responsibility attaches to resolution, not to causal gaps. A resolution can be fully causally explainable and still authored. Determinism describes completed histories; agency occurs at the closure events that constitute those histories. The Displacement Fallacy fails because it assumes authorship is a residual metaphysical ingredient rather than a closure verdict in a control architecture.

4. A Taxonomy: What Competing Theories Are Actually

Doing Many disputes persist because theories answering different questions are treated as competitors.

4.1 Locational theories

Locational theories identify where authorship attaches in action. RT belongs here. It locates attribution at resolution.

4.2 Criterial theories (control and competence accounts)

Criterial theories specify when attribution is fair or warranted. Reasons-responsiveness frameworks (e.g., guidance control / responsiveness criteria) are best interpreted as integrity diagnostics: tests of whether the evaluation→resolution pipeline is functioning sufficiently to justify high-confidence attribution. Under RT, such accounts do not ground agency; they inform confidence, mitigation, and

intervention after agency is located.

4.3 Practice theories (reactive attitudes and moral address)

Reactive-attitudes and related accounts describe how blame, excuse, resentment, indignation, forgiveness, and moral address function in social life. Under RT, these practices are not metaphysical foundations; they are social exposure-feedback mechanisms that stabilise coordination and discourage agency-destroying conduct. Resolution Ethics explicitly reinterprets blame as forward-looking risk management and repair rather than desert-based moral accounting.

4.4 Normativity/ethics theories (constructivism/constitutivism and

contractualism) Constitutivist/constructivist approaches (including Korsgaard-style self-constitution moves) aim to ground normativity in agency rather than moral realism. Contractualist approaches (Scanlon-style) ground wrongness in justifiability to others. Under RT, both become downstream constraint layers: evaluative architectures generate normative pressure under exposure (structural feedback), and shared systems impose additional constraints of justification, reciprocity, and coordination. Resolution Ethics I grounds normativity as “compressed exposure feedback” for systems resolving under time and consequence without moral realism or theology.

4.5 Revisionism as governance strategy under uncertainty

Revisionist approaches (“keep what works, revise what fails”) can be interpreted under RT as institutional strategy: how to maintain agency-preserving practices under epistemic uncertainty about metaphysical freedom. RT supplies the architecture that prevents revision from collapsing into either moralism or nihilism.

5. The Hosting Thesis: RT as Architecture, Others as

Modules Once agency is located at resolution, many theories become modular contributions rather than rival foundations: exposure-coupling. operating within evaluation/resolution. resolutions. uncertainty. This hosting move reduces the incentive for theory-conflict while preserving what is useful in each framework.

- Reasons-responsiveness theories become diagnostics of intact resolution.
- Reactive-attitudes theories become social feedback accounts enforcing
- Constitutivist/constructivist theories become internal binding descriptions
- Contractualist theories become shared-justification constraints on permissible
- Revisionist accounts become policy-level governance strategies under

5.1 The no-authorship retreat (without name-dropping)

A common escape from metaphysical free will disputes is to abandon authorship entirely: dissolve responsibility and replace it with management, prediction, or therapy. RT blocks that retreat without metaphysical inflation: Some approaches avoid metaphysical commitments by dissolving authorship altogether. Resolution Theory avoids metaphysics without paying that price: it preserves authorship by locating it structurally at deliberative closure and separating explanation from attribution. This

preserves accountability while allowing maximal explanatory sophistication.

6. The Regress of Resolution (The “Luck” Objection)

A predictable hard objection is that RT relocates the problem rather than solving it: if resolution is a structural event, what causes it to occur at that moment and in that way? If neuroscience can fully explain the timing and character of closure via prior brain states, fatigue, bias, conditioning, and environmental load, then the critic claims the attribution point is “hollow”—just another link in the causal chain. This objection repeats the Displacement Fallacy. The demand “what causes closure to happen at that moment?” is simply a demand for more causal description. Of course closure has causes. RT never denies that. But causal sufficiency does not answer the attribution question. Explanation tells you how the closure happened; authorship tells you whose closure it was. Treating full explanation as if it makes attribution empty assumes that authorship requires a causal gap. RT rejects that assumption. Resolution is not introduced as a metaphysical status or a mysterious extra ingredient. It is a functional event in a control architecture: the point at which deliberation is closed and a commitment is formed—where possibilities are no longer merely represented and weighed, but taken up as the action the system will do. Neuroscience can, in principle, explain why that closure occurred then rather than later and why it took that shape rather than another. That is consistent with RT. Explanation can become arbitrarily detailed without ever summing into authorship, because authorship is not a causal remainder; it is a verdict about an integrated system’s closure. The critic’s “your resolution-module did it, not you” move is decomposition masquerading as refutation. If the resolution-module isn’t me, then nothing in my body is me—so your objection dissolves the agent by mere decomposition. One can always zoom in: it wasn’t the person, it was the brain; it wasn’t the brain, it was neurons; it wasn’t neurons, it was molecules. That is not an argument against authorship. It is the trivial fact that complex systems are implemented by parts. Agents are integrated systems that evaluate, resolve, persist, and remain exposed to the consequences of their resolutions over time. Authorship attaches to that system’s closure events, not to an uncaused spark. The locus is therefore not hollow. It is not a metaphysical remainder. It is the point in the architecture where commitment happens—where evaluation becomes owned action.

7. Demonstration Cases: Where the Interface Does Real

Work

7.1 Addiction: distortion without bypass

Addiction is a paradigmatic site for the Displacement Fallacy. As explanation becomes dense (reward circuitry, conditioning, withdrawal dynamics), the temptation is to treat the agent as bypassed and authorship as absent. The resolutionist account distinguishes distortion from bypass. Addiction typically distorts evaluation—certain considerations dominate salience and urgency—but does not bypass resolution. Addicted agents plan, conceal, time behaviour, negotiate constraints, and respond to incentives. These behaviours require deliberative closure and therefore authored resolution. The correct consequence is structural: This dissolves the moralism/medicalisation deadlock without cruelty or nihilism.

- responsibility remains (resolution occurred),
- blame is mitigated (evaluation was warped),

- intervention aims at restoring evaluative balance rather than denying agency.

7.2 Institutional harm: agency laundering

In complex organisations, harms emerge through committees, procedures, diffusion, and incentive structures. Explanation proliferates and authorship becomes illegible. This modern pathology is agency laundering: power remains real while responsibility disappears into “the system”. RT forces attribution to land at resolution points in the authorisation chain: sponsorship, sign-off, deployment decisions, and governance choices. Complexity may diffuse causal contribution, but it does not erase closure events. The Displacement Fallacy becomes systemically dangerous here: the more explainable a harm becomes as “structural,” the easier it becomes to evade authorship.

7.3 AI governance: evaluation without authorship

In AI debates, optimisation and evaluation are often mistaken for agency. Systems select outputs and policies, and pressure grows to assign them responsibility. RT blocks this by separating optimisation from authored closure under exposure: responsibility remains with human sponsors. Under subjecthood uncertainty, RT proposes staged provisional protections without premature personhood, while retaining human authorship upstream and imposing disclosure obligations where credible indicators appear. This shows RT functioning as an interface: explain AI behaviour fully, but do not displace human authorship.

8. Diminished Capacity, Bypass, and “Non-Authored

Conduct” (Legal Mapping) The Displacement Fallacy is especially corrosive in legal contexts, where increasing explanation of impairment is routinely misread as excusing authorship. RT blocks this by separating process categories from verdict categories and by distinguishing strict bypass from threshold-based non-authorship.

8.1 Bypass (process category; must remain rare)

Bypass denotes cases in which resolution does not occur at all: the evaluative–deliberative architecture is excluded or replaced, and the behaviour is not authored in any ordinary sense. Paradigm cases include seizure, unconsciousness, and genuine automatism. RT treats bypass as a strict category because over-expanding it collapses responsibility into explanation and licenses agency laundering in law.

8.2 Distortion (common; authorship intact with mitigation)

Most impairment cases are not bypass. They involve distorted inputs and warped evaluation under heavy load: addiction, fear, coercive control, trauma states, intoxication, fatigue, and many forms of mental disorder. In these cases, resolution typically still occurs; deliberation closes into action, but under abnormal conditions. Responsibility remains binary (authored resolution occurred); blame and sanction are adjusted scalar and forward-looking.

8.3 Non-authored conduct (legal threshold category; extreme

distortion) Legal systems require binary verdicts. Psychology supplies gradation. RT therefore distinguishes bypass as a process category from a separate legal threshold category: Bypass is the absence of resolution and must remain rare. “Non-authored conduct” is a legal threshold category:

distortion can become so severe that, although some minimal closure may occur, the act falls below the authorship standard required for criminal guilt and is therefore treated as non-authored for legal purposes. This category is not a metaphysical claim that no closure occurred; it is a legal verdict that the closure, if present, was below the threshold of authorship required for ordinary criminal attribution. This preserves conceptual hygiene (bypass is not inflated), preserves mitigation realism (capacity degrades), and prevents the Displacement Fallacy from turning scientific explanation into blanket exculpation.

9. Objections and Replies

9.1 “This is just compatibilism”

RT is not a semantic redefinition of freedom. It is a separation of verdict-types: explanation vs attribution. The free will debate becomes non-rivalrous once authorship is located at resolution rather than causal gaps.

9.2 “Resolution comes in degrees”

Psychological capacities degrade by degrees; governance requires binary attribution with graded confidence. RT accommodates graded confidence while preserving binary responsibility and scalar blame, and it distinguishes strict bypass from threshold-based legal non-authorship.

9.3 “This derives ought from is”

RT does not infer moral facts from brute description. It identifies structural feedback constraints for systems that resolve under exposure across time. Normativity is navigational: “ought” is compressed exposure feedback, not metaphysical command.

9.4 “This reframes rather than answers”

Reframing is necessary when debates persist due to category confusion. RT preserves the concerns of existing frameworks while clarifying where their claims properly attach.

10. What Resolution Theory Is Not

RT does not deny or affirm metaphysical theses about souls, ultimate reality, or moral realism. It is structurally agnostic. It does not claim agents are uncaused, nor that explanation is morally irrelevant. It does not reduce ethics to preference or culture. It does not imply that blame must be punitive; it rejects desert-based punishment and defends forward-looking containment and repair where required.

11. Conclusion

The Displacement Fallacy—the belief that increasing explanation shrinks agency—rests on a category confusion. Causal explanation operates over inputs and evaluation; agential attribution attaches at resolution. Causal explanation describes the history of closure; attribution identifies its locus. They are different verdict-types regarding the same event. Once this interface is fixed, much of the responsibility literature becomes modular: reasons-responsiveness accounts diagnose intact resolution, reactive attitudes describe exposure-feedback practices, constructivist and contractualist ethics supply downstream constraint layers, and revisionist approaches become governance strategies under uncertainty. In legal contexts, the same interface stabilises diminished capacity doctrine by keeping

bypass rare, treating most impairment as distortion-with-mitigation, and reserving “non-authored conduct” for extreme distortion cases that fall below the authorship threshold required for criminal guilt. The consequence is a responsibility architecture that remains stable under determinism and increasing causal explanation while scaling to institutions, law, and AI without laundering authorship. Explanation expands; authorship attaches at closure. The two do not compete.

References (Internal Corpus)

Easton, H. R. Distinction Without Escape: Resolution, Exposure, and the Location of Free Will.
Easton, H. R. Resolution Theory and the Problem of Talking Past Each Other: Agency, Explanation, and a Recurrent Category Error. Easton, H. R. Distortion Without Bypass: Addiction, Agency, and Responsibility on a Resolutionist Account. Easton, H. R. Resolution Ethics I: Normativity as Structural Feedback. Easton, H. R. Resolution Ethics: Blame Without Desert—Sanction, Containment, and Repair. Easton, H. R. Moral Agency, Emergence Uncertainty, and Civilisational Risk.