Irreversible Commitment as the Basis of Preference

Riaan de Beer
predictiverendezvous@proton.me
Independent Researcher
ORCID: 0009-0006-1155-027X
DOI: 10.5281/zenodo.18601902
https://github.com/infinityabundance/irreversible-commitment
Version 1.0

Abstract

This work introduces a foundational account of preference grounded in irreversible time. Rather than treating preference as valuation, ranking, desire, or utility, it is modeled as a form of *irreversible self-binding* required to preserve directional coherence across time. Systems that cannot rewind or fully revoke prior commitments must bind themselves to future courses of action by permanently excluding incompatible alternatives. On this account, preference arises not from comparative assessment, but from the necessity of committing under conditions of temporal irreversibility. Preference is thus identified with the structural act of reducing future optionality in order to maintain coherent continuation. This reframing positions preference as a commitment-based invariant generated by irreversible time, independent of representation, optimization, or subjective experience.

Contents

1 Introduction

Preference is commonly treated as a matter of valuation: a ranking of options, an assignment of utility, or an expression of desire. Such accounts presuppose that choices can be revised, re-evaluated, or rescored without structural consequence. Under irreversible time, however, this presupposition fails. Once actions are taken and states are entered, alternatives are permanently excluded, and future courses are constrained by past commitments. Any account of preference that ignores this asymmetry cannot explain how direction is preserved across time.

This work begins from a more primitive constraint: systems that persist under irreversible time must commit to future courses of action without the possibility of full revocation. Even in the absence of error, noise, or external pressure, the inability to rewind imposes a requirement to bind future behavior in advance. Preference, in this setting, cannot be reduced to assessment or comparison alone; it must function as a mechanism that renders some futures inaccessible in order to make others possible.

The central claim advanced here is that preference arises as *irreversible self-binding*. To prefer is not merely to judge or rank, but to commit—to exclude incompatible alternatives such that future action remains coherent rather than indeterminate. This exclusion is not incidental; it is the defining feature that distinguishes genuine preference from provisional inclination. Where optionality remains fully recoverable, preference remains unstable.

This account treats preference as a structural invariant generated by irreversible time. It does not depend on representations of value, internal reward signals, or subjective experience. Nor does it appeal to ethical norms or external enforcement. Instead, preference is identified with the act of reducing future optionality in order to preserve directional coherence across temporal extension.

The contribution of this work is therefore not a theory of desire or motivation, but a foundational reframing of preference as a commitment-based phenomenon. By locating preference in irreversible self-binding rather than valuation, it isolates a necessary condition for directional persistence in any system that must continue forward without rewind.

The remainder of the paper proceeds as follows. Section 2 examines how irreversible time necessitates commitment and why reversible choice cannot ground preference. Section 3 distinguishes preference from valuation and clarifies the role of optionality reduction. Section 4 formalizes irreversible self-binding as the primitive underlying preference. Section 5 analyzes stability and failure modes, and Section 6 delineates the scope and boundaries of the account.

2 Irreversible Time and the Necessity of Commitment

Irreversible time imposes a constraint that is independent of interpretation, implementation, or substrate: once a transition occurs, it cannot be undone. Past states cannot be re-entered, and excluded alternatives cannot be restored. This asymmetry is not a contingent feature of particular systems, but a structural condition that governs persistence across temporal extension.

Under reversible conditions, choice can be treated as provisional. Options may be evaluated, rescored, or abandoned without permanent consequence. Preference, in such a setting, reduces to comparison: an ordering that can be revised whenever additional information becomes available. Irreversible time disrupts this model. When actions permanently constrain future possibilities, comparison alone is insufficient to sustain direction. Some mechanism must bind future action despite the impossibility of full revocation.

This necessity gives rise to commitment. Commitment is not the outcome of valuation, but a forward-binding constraint that renders certain futures inaccessible in order to stabilize others. Where time is irreversible, the absence of commitment leads to indeterminacy: unresolved alternatives accumulate, and future action loses coherence. Commitment resolves this by enforcing exclusion in advance of outcome.

Importantly, commitment is required even in idealized conditions. The necessity does not arise from uncertainty, error, or limited rationality. Even a perfectly informed system must commit under irreversible time, because waiting to decide preserves optionality at the cost of direction. Directional persistence requires that some alternatives be abandoned before their outcomes are fully known.

This reveals a limitation of accounts that equate preference with ranking, scoring, or utility maximization. Such constructs presuppose that alternatives remain available for reconsideration. Under irreversible time, however, the defining feature of preference is not the ordering of options, but the act of binding future behavior by permanently excluding incompatible courses.

Preference therefore cannot be grounded in reversible choice. It must instead be grounded in commitment: a structural response to temporal asymmetry that enables continued action without collapse into indecision. The next section develops this distinction by separating preference from valuation and identifying optionality reduction as its defining characteristic.

3 Preference Without Valuation

Accounts of preference are commonly grounded in valuation. Preferences are treated as expressions of desire, assignments of utility, or rankings among alternatives. On such views, to prefer is to assess—to compare options according to some internal or external metric. While this framing is widespread, it presupposes that preference remains revisable without structural consequence. Under irreversible time, this presupposition does not hold.

Valuation is intrinsically comparative. It operates by ordering alternatives while leaving them, in principle, available for reconsideration. A ranking can be updated; a score can be revised; a preference ordering can be inverted. None of these operations, by themselves, require the permanent exclusion of alternatives. As a result, valuation alone cannot account for how direction is preserved once alternatives must be irreversibly abandoned.

Preference, as treated here, is distinct from valuation. To prefer is not merely to judge that one option is better than another, but to bind future action by rendering certain possibilities inaccessible. This binding is not an accidental byproduct of choice; it is the defining feature of preference under irreversible time. Where no such binding occurs, what is present is assessment without commitment, not preference proper.

The defining characteristic of preference is therefore the reduction of future optionality. To establish preference, a system must incur a loss: the permanent exclusion of incompatible futures. This loss is not a failure or limitation, but the mechanism by which coherent continuation is achieved. Without optionality reduction, direction dissolves into perpetual deferral.

This reduction must also be costly in a structural sense. If excluded alternatives can be trivially reinstated, commitment loses its binding force and preference becomes unstable. Genuine preference requires that reversal be non-trivial or impossible, such that the system's future is meaningfully shaped by its prior commitments.

Preference without valuation is thus not an absence of assessment, but a reordering of explanatory priority. Valuation may inform which commitments are made, but it does not constitute preference

itself. Preference arises only when assessment is converted into irreversible self-binding, and it is this conversion—rather than comparison—that grounds preference under irreversible time.

The following section formalizes this process by defining irreversible self-binding and identifying it as the primitive underlying preference persistence.

4 Irreversible Self-Binding

The preceding sections motivate the need for a mechanism that converts assessment into durable direction under irreversible time. This section formalizes that mechanism as *irreversible self-binding* and identifies it as the primitive underlying preference.

Irreversible self-binding is the act by which a system commits itself to a future course by permanently excluding incompatible alternatives. This act is not defined by the content of what is chosen, but by the structural consequences of choosing. A self-binding is irreversible when the exclusion it effects cannot be undone without violating the system's continued coherence across time.

Self-binding differs from decision in that it is forward-binding rather than evaluative. A decision may select among options at a moment; self-binding constrains what remains possible thereafter. The binding operates not by storing a choice, but by shaping the space of future action such that certain paths are no longer available. In this sense, self-binding is not a state but a transformation of possibility.

The irreversibility of self-binding is essential. If excluded alternatives can be reinstated without cost, the binding fails to stabilize direction. Preference persistence requires that reversal be structurally non-trivial or impossible, so that future behavior is genuinely conditioned by past commitment. Irreversibility need not imply immutability of internal structure, but it does require that incompatible futures cannot be simultaneously preserved.

Irreversible self-binding is also temporally anchored. It arises at a particular moment of commitment and derives its force from the impossibility of returning to the pre-commitment configuration. The binding is therefore inseparable from the temporal context in which it is enacted. Its authority does not stem from justification or endorsement, but from the fact that the system has already passed the point at which alternatives could coexist.

Within this framework, preference is identified with irreversible self-binding. To prefer is to bind oneself such that future action is constrained by the permanent exclusion of alternatives. Valuation, deliberation, and comparison may precede this act, but they do not constitute preference until self-binding occurs. Preference persistence across time is therefore a consequence of the durability of self-binding, not of the stability of evaluative judgments.

The next section examines how failures of irreversible self-binding lead to preference instability, oscillation, and loss of directional coherence, and delineates the structural limits on preference revision under irreversible time.

5 Preference Stability and Failure Modes

Irreversible self-binding provides the structural basis for preference persistence under irreversible time. When such binding is absent, incomplete, or trivially reversible, preference loses its stabilizing function. This section examines the characteristic failure modes that arise when commitment fails to bind future action, and clarifies the limits within which preference revision remains compatible with directional coherence.

The primary failure mode is instability under reversible commitment. When exclusions are provisional or easily revoked, alternatives remain effectively available, and preference oscillates rather than persists. Direction becomes contingent on momentary assessment, leading to repeated re-evaluation without convergence. Under irreversible time, such oscillation is not neutral; it erodes the system's capacity to act coherently, as future courses are continually reopened rather than consolidated.

A related failure mode is indecision through accumulation. When commitments are deferred in order to preserve optionality, unresolved alternatives compound across time. The system retains the appearance of flexibility at the cost of direction. As irreversible transitions continue, this accumulation eventually forces exclusion under degraded conditions, resulting in commitments that are no longer guided by coherent preference but by structural necessity alone.

Fragmentation constitutes a more severe failure. In this case, incompatible commitments coexist without effective exclusion, leading to internal conflict over future direction. Because irreversible time prevents simultaneous realization of incompatible futures, such fragmentation undermines preference persistence by preventing any single commitment from binding future action. Directional coherence is lost, not because commitments are absent, but because none attains sufficient exclusivity to dominate.

These failure modes illustrate that preference stability depends not on the correctness or optimality of commitments, but on their binding force. Preference can be revised, but only within limits imposed by prior exclusions. Revision is structurally possible when it preserves coherence with existing commitments; it becomes destabilizing when it attempts to reinstate alternatives that irreversible self-binding has already excluded.

Preference revision under irreversible time therefore differs fundamentally from reassessment under reversible conditions. It is not a return to a neutral evaluative state, but a constrained transformation that must respect the commitments already enacted. The cost of revision reflects the degree to which prior exclusions must be renegotiated, and in extreme cases, revision entails the abandonment of identity or direction altogether.

These considerations delineate the structural role of irreversible self-binding in maintaining preference stability. Where commitment binds effectively, preference persists despite uncertainty and change. Where binding fails, preference collapses into oscillation, indecision, or fragmentation. The final section situates this account within its intended scope and clarifies its boundaries.

6 Scope and Implications

The account developed here is intentionally minimal. It does not propose a theory of desire, motivation, or evaluation, nor does it appeal to psychological, ethical, or computational mechanisms. Its aim is to isolate a structural condition imposed by irreversible time and to identify the form of commitment required for preference to persist across temporal extension.

The analysis is substrate-independent. Irreversible self-binding is defined in terms of structural exclusion of alternatives, not in terms of particular physical, biological, or representational realizations. Any system that persists under irreversible time and must maintain directional coherence is subject to the same constraint, regardless of how its internal states are implemented or interpreted.

This account is also distinct from normative or ethical theories of preference. It does not address what ought to be preferred, nor does it evaluate preferences in terms of justification, rationality, or moral standing. Preference, as treated here, is prior to normativity: it concerns the binding of future

action, not the reasons for doing so. Normative considerations may guide which commitments are formed, but they do not constitute preference itself.

The relationship between preference and identity is likewise indirect. Irreversible self-binding contributes to directional persistence, while identity persistence concerns continuity of self-reference across change. Although commitments may shape identity over time, the present account does not attempt to define identity or to specify the conditions under which it is preserved. It isolates only the commitment structure required for preference stability.

Finally, this framework does not deny the role of valuation, deliberation, or comparison. Such processes may inform commitment and influence which alternatives are excluded. However, they are insufficient to ground preference in the absence of irreversible self-binding. The scope of this work is therefore limited to identifying the structural act that converts assessment into binding direction under irreversible time.

The following conclusion summarizes the core claim and reiterates the foundational contribution of identifying preference with irreversible commitment rather than valuation.

7 Conclusion

This work has advanced a foundational account of preference grounded in irreversible time. By rejecting valuation, ranking, and utility as sufficient explanations, it has identified preference with a structural act of irreversible self-binding. Under conditions where time cannot be rewound and excluded alternatives cannot be recovered, preference must function to bind future action by permanently reducing optionality.

Irreversible self-binding was introduced as the primitive underlying this function. Commitment, on this account, is not a derivative of assessment but a necessary response to temporal asymmetry. Preference persists not because evaluations remain stable, but because commitments enact exclusions that cannot be trivially undone. Directional coherence across time is thus secured by the durability of binding rather than by the consistency of valuation.

The analysis further clarified the limits of preference stability and the characteristic failure modes that arise when binding is reversible, deferred, or fragmented. Oscillation, indecision, and loss of direction follow when commitment fails to exclude alternatives with sufficient force. Preference revision remains possible, but only insofar as it respects the exclusions already enacted under irreversible time.

The central contribution of this work is the identification of preference as a commitment-based invariant generated by temporal irreversibility. By locating preference in irreversible self-binding rather than in evaluative comparison, it isolates a necessary structural condition for directional persistence in any system that must continue forward without rewind. This reframing establishes preference as a consequence of temporal constraint, independent of representation, optimization, or subjective experience.