Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identifying Commitment-Gated Choice Reversal:Exact Limits and Design Requirements

Daniel Bryan

danielbryanldn@gmail.com

Research Article

Keywords: choice reversal, commitment, model identification, decision dynamics, switching costs

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11412694/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Identifying Commitment-Gated Choice Reversal: Exact Limits and Design Requirements

Abstract

When does prior commitment make a choice genuinely harder to reverse rather than simply less attractive? This article develops Structural Revisability Theory (SRT) as a restricted process account of history-dependent choice reversal and then asks whether its distinctive commitment gate is identifiable. Three results follow. First, an exact identification limit: in the reduced gated model $U = \alpha + A[1 - \exp\{-\rho dq(1 - K_0)\}]$, fixing initial accessibility K_0 makes the reparameterization $\rho' = \rho(1 - K_0)$ reproduce a generic saturating rival at every dwell and activation level. No sample size breaks that equivalence. Second, in an equal-complexity synthetic binary-choice comparison, merely crossing two accessibility levels yields weak discrimination: across 800 synthetic datasets, pooled correct selection is 63.8% (95% CI 60.4–67.0). Third, a one-factor-at-a-time decomposition identifies the useful design variation. Widening initial accessibility alone raises recovery to 83.9% (81.2–86.3), whereas longer dwell adds only 0.3 percentage points. The result is methodological and theoretical rather than behavioral evidence: tests of commitment-gated reversal must vary independently measured initial reversibility, not simply observe persistence for longer.

Keywords: choice reversal; commitment; model identification; decision dynamics; switching costs

Identifying Commitment-Gated Choice Reversal: Exact Limits and Design Requirements

1. Choice Reversal After Commitment

People often retain an incumbent choice after new information makes an alternative more attractive. Persistence can reflect conservative or motivated updating, reference dependence, sunk-cost sensitivity, status-quo preference, switching costs, decision noise, or rational continuation. SRT does not treat persistence itself as evidence for a new mechanism. It asks a narrower question: can active occupation of a difficult-to-reverse choice state alter the support later required for information-responsive reversal?

The strongest claim concerns state history rather than calendar time. Two decision makers can begin with comparable beliefs, incentives, and choices yet later differ in reversibility because one has repeatedly reactivated a low-revisability commitment. The target outcome is renewed sensitivity of choice to diagnostic information, not switching frequency by itself.

This distinction is relevant across project continuation, consumer switching, portfolio reallocation, policy choice, and other decisions in which an incumbent state carries monetary, procedural, reputational, or self-consistency consequences. SRT separates information registration, conditional selection, and transition feasibility, and requires the last construct to be measured or manipulated independently rather than inferred from non-reversal.

Figure 1 summarizes this explanatory hierarchy and the nested SRT-0 to SRT-2 claim ladder. SRT-0: constrained reachability is a useful representation of economically consequential belief-choice revision. SRT-1: target-specific transition variables add prospective prediction beyond updating responsiveness, motivated beliefs, sunk-cost exposure, and ordinary switching costs. SRT-2: active low-revisability occupancy changes future reopening requirements after present state and incentives are matched. The strongest novelty resides in SRT-2.

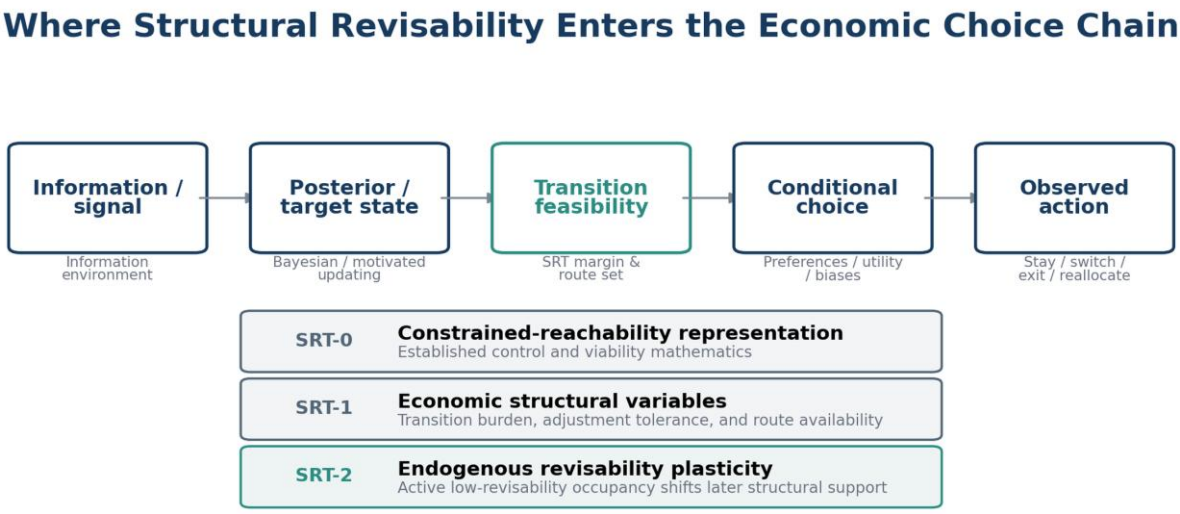


Figure 1. Explanatory hierarchy and claim ladder. Established judgment-and-decision accounts already address information processing, belief formation, preference-dependent selection, and status-quo persistence. SRT locates its additional claim at transition feasibility and its principal novelty at SRT-2: endogenous modification of later reopening requirements by active low-revisability occupancy.

1.1 Revision as a Constrained Decision Transition

Let s_t denote the current belief-choice state and Ω_t the relevant state space at time t. The state can include posterior beliefs, the current action, objective commitments, and psychologically consequential features of prior choice. For a candidate revision ϕ, define $R\phi(t)$ as the set of states satisfying a preregistered criterion for successful incorporation of new information, such as an evidence-responsive posterior together with a corresponding change in action.

The existence of an attractive alternative does not imply that the system can reach it under current constraints. Changing course can impose transaction costs, regret, accountability, coordination burdens, or discontinuities created by the option set. The appropriate object is therefore a directed path through a belief-choice state space rather than a static comparison between endpoints.

Let $\Pi H(s_t, R\phi)$ denote admissible transition paths from the current state to the revision-consistent target within horizon H. The horizon is operational: a choice that cannot be reversed within a laboratory round may become reversible after a cooling-off period, a reduction in transaction costs, a partial-adjustment option, or a change in accountability.

A path π carries a transition burden $J_t(\pi)$. In the scalar reduction, the minimum currently available transition burden is

$$C_r^*(s_t, \varphi; H) = \inf \{ J_t(\pi) : \pi \in \Pi_H(s_t, R_\varphi) \}$$

Adjustment tolerance T_t denotes how much transition burden the current configuration can absorb while still completing the change. Structural revisability is then represented by

$$S_H(s_t, \varphi) = 1 \iff C_r^*(s_t, \varphi; H) \leq T_t$$

The scalar inequality is a feasibility representation, not a new psychological law. Its equivalence to an actual feasible path assumes attainment of the infimum under suitable regularity conditions; the existential path condition in Section 3.2 is primary. Empirical content begins only when the relevant constraints are measured independently and predict later reversal.

1.2 The Revision Capture Basin

The reachability formulation has a direct formal antecedent in viability theory, which studies dynamical systems under state and control constraints and defines capture basins for target states (Aubin, 1991, 2001). SRT therefore does not claim to invent constrained target reachability. Its use here is to represent economically consequential revision as a path-dependent feasibility problem.

Define the revision capture basin over horizon H as

$$C_H(R_\varphi) = \{ s : \exists \text{ admissible } \pi \text{ from } s \text{ to } R_\varphi \text{ satisfying the revision constraints} \}$$

Then revision is structurally reachable when

$$s_t \in C_H(R_\varphi)$$

The decision-theoretic question is what determines the boundary of this capture basin and whether that boundary remains fixed. SRT proposes that constraints can be target-specific, include both objective and psychological transition burdens, and potentially change through active occupation of a low-revisability belief-choice state. Figure 2 illustrates the resulting revision landscape.

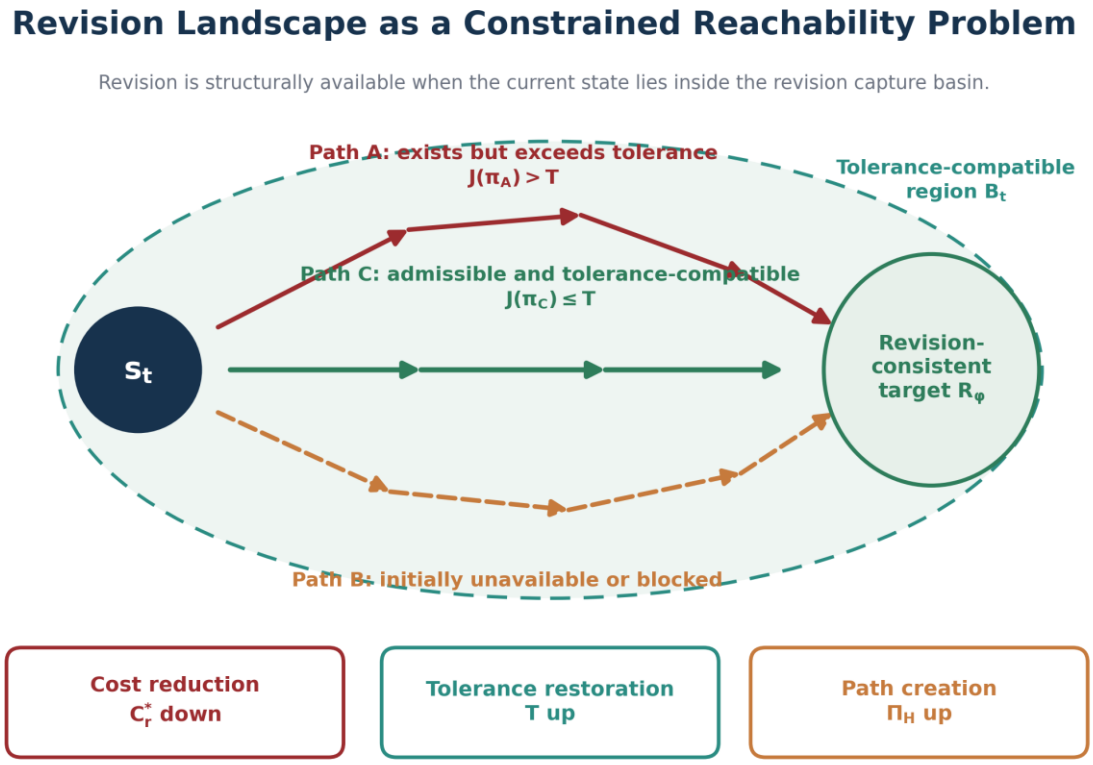


Figure 2. Revision landscape as constrained choice reachability. Multiple routes can connect a current belief-choice state to a revision-consistent target. Reopening can occur by lowering the burden of an existing route, increasing adjustment tolerance, or creating an intermediate route such as partial adjustment, delayed commitment, or reversible switching.

1.3 Updating Bias, Choice Resistance, and Structural Unavailability

Observed persistence can arise at several levels. Information may be misunderstood or underweighted; a correctly updated belief may still favor the incumbent once payoffs and preferences

are represented; or the required transition may be unavailable under current constraints. The same observed choice can therefore reflect different mechanisms.

A useful factorization is

$$P(Y_{rev} = 1) = K_t \times Q_t(\varphi \mid E, \vartheta)$$

Here K_t is effective transition accessibility and Q_t is conditional selection given accessibility. The factorization does not identify either quantity from choice alone: many pairs of K and Q yield the same observed probability. Conditional selection must therefore absorb ordinary valuation, risk preference, status-quo preference, switching cost, and decision noise.

Observed non-revision cannot by itself diagnose structural closure. Any empirical test must therefore measure signal comprehension, posterior movement, current incentives, and ordinary switching costs separately from the later reopening outcome.

1.4 The Residual Novelty Claim

The distinctive empirical commitment is not status-quo bias, sunk-cost persistence, switching costs, network entrenchment, thresholding, or ordinary hysteresis. It is that active occupation of a low-revisability belief-choice configuration can modify the future revision landscape. Let q(t) denote activation of the relevant commitment configuration and K(t) current revision accessibility. Define active low-revisability occupancy as

$$z(t) = q(t)[1 - K(t)]$$

The core dynamic hypothesis is

$$\mathcal{L}(t + \Delta) = F[\mathcal{L}(t), z_t, E_t, X_t, u_t]$$

where $\mathcal{L}(t)$ denotes the revision landscape, E_t the information environment, X_t exogenous context and payoff conditions, and u_t an intervention or choice-architecture change. Endogenous plasticity exists only if future landscape parameters depend on occupancy after present beliefs, incentives, information, and elapsed time are controlled.

Let Ureopen* denote the minimum non-informational support required to restore information-responsive reversal after validated low revisability. Support could take the form of fee relief, reversibility, reduced accountability, or an intermediate route. In a deepening regime, the distinctive prediction is

$$\partial U^{*}_{reopen} / \partial \Lambda_{\rho} > 0$$

where Λρ is cumulative active low-revisability occupancy with possible decay. The relation is a hypothesis for a deepening regime, not a universal law. Dynamic switching-cost, reference-adaptation, reinforcement, and learning models can reproduce similar history dependence; SRT must therefore earn value through independent structural measures and predictive discrimination.

1.5 Contribution and Structure

The completed contribution of this article is an identification result, not evidence that the mechanism governs human choice. Section 9.2 proves that, when initial accessibility is fixed, the reduced commitment-gated reopening curve is exactly observationally equivalent to a generic saturating rival after a rate reparameterization. No increase in sample size can break that equivalence.

Sections 9.3–9.4 then ask what design variation makes the models recoverable against an equal-complexity rival. Crossing two accessibility levels yields only 63.8% pooled correct selection in the base binary-choice design, pooled over 800 synthetic datasets. Widening initial accessibility alone raises recovery to 83.9%, whereas doubling maximum dwell changes it by only 0.3 percentage points. The central design lesson is therefore to vary independently measured starting reversibility rather than simply observe commitment for longer.

The identification problem is embedded in the wider theory because it follows directly from the proposed gate. Sections 3–6 define transition burden, adjustment tolerance, and path availability; Section 7 gives an incentivized choice-reversal design; Section 8 connects the framework to judgment and decision research. The mechanism itself remains untested.

Scope is limited to directed choice revisions with explicit targets; rational persistence and ordinary switching costs remain direct alternatives.

Section 2 reviews incumbent explanations; Sections 3–8 develop the model and test logic; Section 9 establishes identification limits and design requirements; Section 10 states the remaining empirical burden.

2. Decision-Theoretic Antecedents and Rivals

The theory should be assessed against strong incumbent explanations rather than against persistence alone. Belief updating, motivated cognition, sunk costs, status-quo preference, switching costs, reference dependence, nonlinear dynamics, and constrained control already explain substantial parts of choice persistence. Table 1 locates the residual claim after that territory is conceded.

Table 1. Novelty firewall: established explanations and the residual SRT contribution.

Framework	Already provides	SRT must not claim as novel	Residual SRT role
Bayesian / motivated-belief models	Belief updating under priors, signal interpretation, and belief utility	Biased or asymmetric updating	Transition feasibility after information processing is measured
Sunk cost / escalation accounts	Persistence after prior investment, responsibility, and self-justification	Commitment persistence or escalation	Whether active low-revisability occupancy shifts later reopening
Status-quo / switching-cost models	State-dependent choice and costly movement away from current options	Inertia or transition costs	Endogenous change in the future switching/revision boundary
Viability / control theory	Constraint-respecting target reachability and intervention effort	Reachable sets or minimum control effort	Behavioral-economic parameterization and occupancy-dependent deformation
Network / nonlinear models	Connectivity, bistability, discontinuity, and hysteresis	Entrenchment, thresholds, or ordinary hysteresis	Movement of effective controls after matched present state
Choice architecture / option-set design	Defaults, constrained menus, gradual adjustment, and transaction design	Effects of adding or removing options	Path creation as a distinct reopening mechanism

2.1 Belief Updating and Motivated Choice

Confirmatory bias, utility from beliefs, and motivated updating can produce persistent choices without a separate feasibility constraint (Akerlof & Dickens, 1982; Drobner & Goerg, 2024; Rabin &

Schrag, 1999; Zimmermann, 2020). Responsiveness to information also differs consistently across individuals and predicts how much they value information (Ambuehl & Li, 2018); information that was costly to acquire is weighted more heavily than the same information obtained free (Robalo & Sayag, 2018); and induced motivated beliefs can persist through private signals and market trading (Friedman & Wang, 2026). SRT does not compete at the signal-processing stage. Its residual question is whether prior low-revisability occupancy predicts later support requirements after comprehension, beliefs, incentives, and current choice are represented.

2.2 Commitment, Sunk Costs, and Responsibility

Prior investment, responsibility, and self-justification can increase continuation (Arkes & Blumer, 1985), and choice ownership can affect later learning (Bachmann, 2024). These are direct rivals rather than evidence for SRT. The proposed test therefore randomizes subsequent occupancy after baseline commitment and information processing have been recorded.

2.3 Status Quo Dependence and Switching Costs

Status-quo preference and switching-cost models already provide a language for costly movement away from an incumbent (Samuelson & Zeckhauser, 1988; Guney & Richter, 2018). A sufficiently flexible dynamic switching-cost model can encode the proposed history effect. SRT is useful only if its restricted state variables improve prediction beyond that alternative.

2.4 Control, Network, and Nonlinear Antecedents

Viability theory formalizes constraint-respecting target reachability, network control theory quantifies state-transition effort, and psychological network and cusp models already generate stability, thresholding, and hysteresis (Aubin, 1991, 2001; Dalege et al., 2016, 2019; Henry et al., 2022; Stocker et al., 2023; van der Maas et al., 2003). SRT therefore does not claim constrained reachability, minimum transition effort, or hysteresis as novel.

2.5 Choice Architecture and Path Creation

Choice environments differ not only in switching cost but also in available routes. Binary stay-or-exit menus, partial adjustment, cooling-off periods, delegation, and reversible trials can create different transition paths even when terminal outcomes are similar. SRT treats path creation as distinct from reducing the burden of an existing route.

2.6 Closest Substantive Rivals

Reluctance to realize losses relative to gains, the disposition effect, is documented in both field and experimental trading (Odean, 1998; Weber & Camerer, 1998), and stop-loss orders reduce it only partly (Talpsepp & Vaarmets, 2019). Reference-dependent and realization-based accounts show why reluctance to reverse can persist even when information is understood (Barberis & Xiong, 2012; Frydman & Rangel, 2014). Monitoring and reactivation can also change reference points or salience: viewing a portfolio at login creates a new reference point that produces its own disposition effect (Quispe-Torreblanca et al., 2025). A clean occupancy manipulation must therefore avoid new outcome information and forced justification while measuring reference change as a rival process.

2.7 Residual Proposition

The residual proposition is that active occupation changes future transition requirements disproportionately when the incumbent state is initially low in revisability, after baseline state and generic saturation are represented. The commitment gate and the burden-tolerance-path decomposition are separate claims; either can fail.

SRT-1 is unsupported if structural variables add no held-out predictive value beyond established state variables. SRT-2 is unsupported if occupancy does not shift later reopening, or if generic reactivation, elapsed time, reference adaptation, or dynamic switching costs predict the same response equally well.

3. Structural Revisability as a Constrained-Reachability Model

The model distinguishes the current belief-choice state, admissible controls, transition paths, transition burden, and state constraints. Let the state evolve according to

$$ds/d\tau = f[s(\tau), u(\tau), \xi(\tau), \tau]$$

where u(τ) is an admissible internal or external control, such as switching support or option redesign, and ξ(τ) represents disturbances or unmodelled input. Let U[s(τ),τ] denote the admissible control set and K(τ) the state-constraint set. A path is admissible when controls and states satisfy these constraints throughout the horizon.

3.1 Transition Burden as a Functional

A path burden can be written generically as

$$J_t(\pi) = \int_t^{t+H} L[s(\tau), u(\tau), \tau]\, d\tau + \Phi(s(t + H))$$

where L is an instantaneous transition-burden function and Φ an endpoint term. This is a generic optimization representation, not a claim that all psychological or economic frictions can be reduced to monetary utility without empirical validation.

For economic applications, transition burden may be multidimensional. Define

$$c_t(\pi) = [\text{cswitch}(\pi), \text{closs}(\pi), \text{crep}(\pi), \text{cself}(\pi)]^T$$

representing objective switching or transaction costs, realized-loss consequences, reputational or accountability consequences, and self-consistency or responsibility-related burden. The components need not initially be commensurable, and only some are necessarily pecuniary.

3.2 Adjustment Tolerance as a Feasible Region

Let B_t be a downward-closed tolerance region containing burden configurations that the current agent-environment system can absorb while completing the transition. Tolerance can reflect reversibility, slack resources, scope for experimentation, or the ability to absorb uncertainty and regret without reverting. A path is tolerance-compatible when

$$c_t(\pi) \in B_t$$

The most general structural condition is therefore

$$S_H(s_t, \varphi) = 1 \iff \exists \pi \in \Pi_H(s_t, R_\varphi) : c_t(\pi) \in B_t$$

This formulation preserves the distinction between a feasible transition and an attractive one. Psychological burden and tolerance require an identified measurement scale; a choice response alone cannot determine both. Objective constraints should be documented separately from preference costs.

3.3 Scalar Reduction and the Primary Structural Margin

When an empirically adequate monotone aggregation ψ_t exists, define $J_t(\pi) = \psi_t[c_t(\pi)]$ and

$$C_r^* = \inf \{ J_t(\pi) : \pi \in \Pi_H \}$$

Using the optimization convention inf(empty set) = +infinity makes topological closure a limiting case of the same formalism: if no path exists, C_r^* is infinite. Substantively, however, it remains useful to distinguish an absent route from a costly route because the interventions differ.

Define the structural margin

$$M_t = T_t - C_r^*(t)$$

The margin is the primary reduced variable. Positive M indicates slack, values near zero indicate boundary sensitivity, and negative M indicates budget closure. A common rescaling of burden and tolerance preserves the feasibility relation; separate levels are not identified without scale anchors. The no-path case and the possible failure of infimum attainment remain governed by the existential definition.

A probabilistic accessibility mapping may be written

$$K_t = G_{\Pi,t} \cdot g(M_t)$$

where GΠ,t is a path-availability gate and g a monotone link. A logistic g is a convenient observation model, not a theoretical requirement.

3.4 Three Sources of Persistent Choice

The model distinguishes conditional choice resistance from two forms of structural unavailability. In ordinary selection resistance, a tolerable route exists but expected utility, preferences, beliefs, or behavioral biases favor staying. In budget closure, candidate routes exist but their transition burden exceeds current tolerance. In topological closure, the available menu contains no admissible route to the target within the horizon.

Reopening can involve fee relief, additional resources, reversibility or a new implementation route. Labels do not identify mechanisms: rebates also change net utility, and partial exposure or reversibility can change risk and option value. The decomposition requires selective process measures and explicit financial alternatives.

3.5 Revisability Is Choice-Specific and Directional

Revisability should be indexed as Rev(i,ϕ,t,c,H), not treated as a global trait. The same decision maker can readily switch in one market and remain constrained in another. Direction also matters:

$$C^*(A \rightarrow B) \text{ need not equal } C^*(B \rightarrow A)$$

Moving from A to B may realize losses, forfeit a contractual benefit, or require public reversal even when B to A does not. A short-horizon transition may be inaccessible while a longer-horizon staged path exists. These asymmetries are natural consequences of transition-based choice.

3.6 Rationality and Normative Neutrality

Structural revisability is normatively neutral. Persistence can be optimal when switching is genuinely costly, and high revisability can produce noisy or premature switching. Welfare claims require explicit payoff and preference benchmarks.

4. Nonlinear Choice Dynamics Without Rediscovering Hysteresis

A cusp catastrophe provides a useful reduced description when an economic choice variable changes discontinuously as information or incentives vary. Let x denote a one-dimensional choice or commitment coordinate and define the canonical potential

$$V(x; a, b) = x^4/4 - ax^2/2 - bx$$

with gradient dynamics

$$dx/dt = -dV/dx = -x^3 + ax + b$$

Equilibria satisfy $x^3 - a x - b = 0$, and the fold boundary is

$$27b^2 = 4a^3$$

For sufficiently positive a, two stable equilibria can coexist, and sweeping b can generate hysteresis even with fixed a. This established geometry means that thresholding or hysteresis alone is not distinctive evidence for SRT. Cusp models can be fitted to behavioral data with established methods (Grasman et al., 2009).

4.1 Cusp Dynamics as a Reduced Projection

The observed choice coordinate x is not revision accessibility K, and the cusp splitting parameter a is not identical to transition burden. The effective cusp controls can be treated as reduced functions of the richer revision landscape; the cusp is therefore an alternative projection used to distinguish fixed-landscape from plastic hysteresis, not the ontology of the theory.

4.2 Static Hysteresis

For fixed $a > 0$, the folds occur at

$$b\pm = \pm [2/(3\sqrt{3})]\, a^{3/2}$$

Forward and reverse transition points differ even when the landscape is fixed. Static hysteresis therefore demonstrates path dependence within a landscape, not endogenous modification of that landscape. Figure 3 contrasts this static case with the plastic case developed next.

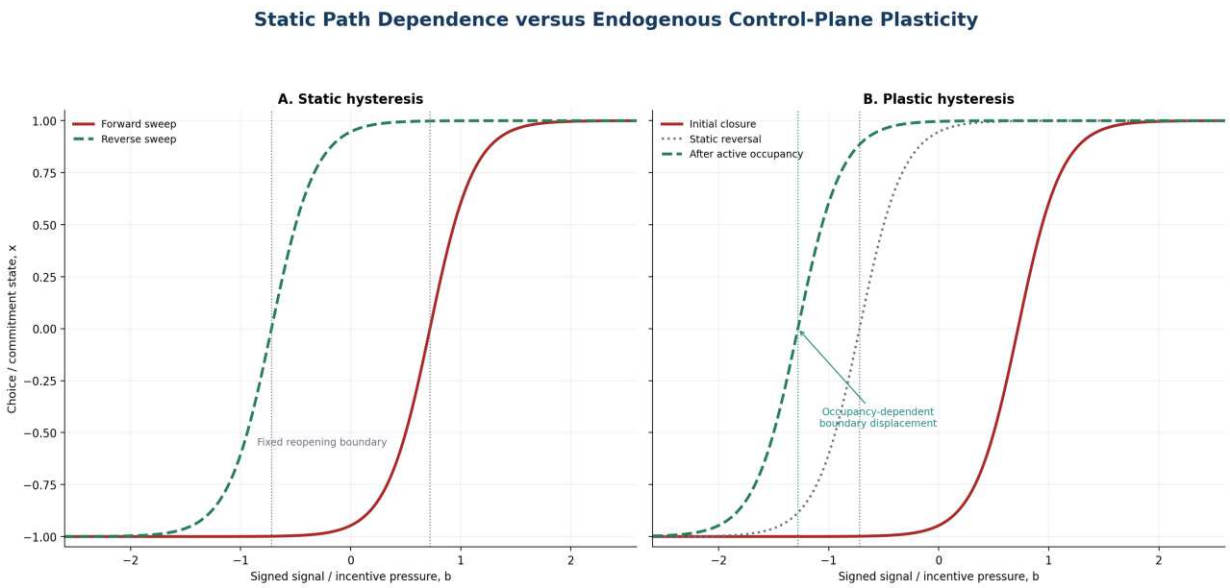


Figure 3. Static versus plastic hysteresis. Ordinary hysteresis can arise within a fixed nonlinear choice landscape. SRT requires the stronger result that active low-revisability occupancy displaces the later reopening boundary because effective transition constraints have changed.

4.3 Plastic Hysteresis

A minimal reduced extension allows the splitting parameter to evolve with active low-revisability occupancy:

$$da/dt = \mu z(t) - \rho[a(t) - a_0] - \eta u(t)$$

The distinctive reduced-form parameter is μ. If μ = 0 and a begins at baseline, occupancy has no independent effect on the future control plane; if μ > 0 under controlled low-revisability occupancy, the effective geometry changes as a function of state history.

The constant-z expression is local because accessibility generally changes as the landscape evolves; confirmatory fitting should therefore integrate the coupled system numerically.

The robust prediction is occupancy-dependent displacement of the reopening boundary, not any universal curvature of the observed choice function.

4.4 Structural Reopening Under a Fixed Information Battery

A clean test holds the post-intervention information battery constant across structural-support conditions. For a fixed signed information context b_r, the cusp boundary implies a critical splitting value

$$a_{crit}(b_r) = [27b_r^2/4]^{1/3}$$

On the appropriate stable branch and with information directed toward the target, suppose support locally reduces the splitting parameter as $a^+ = a^- - \eta u$, with $\eta > 0$. The quasistatic fold-crossing dose is

$$u^*(d) = [a(d) - a_{crit}(b_r)]_+ / \eta$$

This is a local deterministic expression, not a universal behavioral dose law. Finite-horizon dynamics, noise, path availability, and bounded support can prevent completion after fold crossing.

5. Endogenous Revisability Dynamics After Commitment

The reduced cusp identifies what landscape plasticity would look like. The behavioral model must specify which transition constraints change beneath that projection.

5.1 Active Occupancy Rather Than Calendar Time

Define

$$z(t) = q(t)[1 - K(t)]$$

and cumulative active occupancy with exponential decay

$$\Lambda_\rho(t) = \int_0^t \exp[-\rho(t - \tau)]\, q(\tau)[1 - K(\tau)]\, d\tau$$

Calendar time and active occupancy are deliberately separated. A dormant choice held for an hour is not assumed to have the same effect as a repeatedly reactivated choice held over the same interval.

5.2 Transition-Burden and Adjustment-Tolerance Dynamics

A minimal scalar decomposition is

$$dC_r/dt = \beta_E E(t)z(t) + \beta_Q z(t) - \lambda_C[C_r(t) - C_r0] - u_C(t)$$

$$dT/dt = -\gamma_Q z(t) + \lambda_T[T_0 - T(t)] + u_T(t)$$

βE represents discrepancy-contingent burden amplification; βQ occupancy-associated reinforcement; γQ occupancy-associated tolerance erosion; λC and λT recovery; and uC and uT targeted support. These are local empirical parameters rather than universal psychological constants.

These equations deliberately allow recovery. Closure is not assumed to deepen automatically. Define M = T - C_r. Then

$$dM/dt = -z(\gamma_Q + \beta_Q + \beta_{EE}) + \lambda_T(T_0 - T) + \lambda_C(C_r - C_r0) + u_T + u_C$$

The sign of dM/dt distinguishes closed-but-recovering, metastable, and closed-and-deepening regimes.

5.3 Recovery Need Not Follow One Exponential

With different burden and tolerance recovery rates, the raw closure deficit is a difference of exponentials. The general theory therefore does not require a single saturation curve; simpler curves are phenomenological rivals rather than mathematical consequences.

5.4 Structural Memory in Choice

Persistent deviations in burden, tolerance, or path availability constitute structural memory: agents with the same current belief and action can require different later switching support. Current choice need not be a sufficient statistic for revisability.

5.5 Path-Topology Plasticity and Option Design

A path-creating intervention can reopen revision even if the burden on existing routes is unchanged. Partial adjustment, delegation, delayed execution, or a reversible trial can create an intermediate route that a binary stay-or-switch menu does not contain.

5.6 Reopening Support

For budget closure, define $D = [C_r - T]_+$. If a cost intervention has efficacy η_C and a tolerance intervention efficacy η_T, boundary crossing requires approximately

$$\eta_{CuC} + \eta_{TuT} \geq D$$

Equivalent behavioral reopening can arise from different latent operations, so behavior alone cannot identify burden versus tolerance.

A stronger mechanism claim requires differential intermediate effects: a burden-targeting treatment should affect independently measured burden more than tolerance, and vice versa. Both can improve the common margin.

In decision experiments, a subsidy may lower transition burden, a reversible option may increase adjustment tolerance, and menu redesign may create a route. Observed switching alone cannot identify which mechanism changed.

6. Measurement and Identification

SRT is empirically empty if structural variables are inferred retrospectively from failure to switch. Transition burden, adjustment tolerance, path availability, and active occupancy must be measured or manipulated independently of the critical reopening outcome.

6.1 Margin First, Decomposition Second

The primary confirmatory target should be structural margin or capture-basin membership. Separate burden and tolerance factors are a stronger mechanistic layer, not a prerequisite for testing occupancy-dependent plasticity.

6.2 Transition Burden

Candidate indicators include objective switching costs, anticipated regret, accountability, coordination costs, and self-consistency consequences. The relevant quantity is the burden of a directed transition, not total prior investment or a generic dislike of changing one's mind.

6.3 Adjustment Tolerance

Tolerance is transition-specific capacity to absorb the uncertainty or consequences of changing course. It can be affected by slack, reversibility, error-correction opportunities, or the possibility of testing an alternative before an irreversible switch.

6.4 Path Availability

Path availability concerns the routes permitted by the choice architecture. A route counts only if it is feasible under the task rules and can plausibly deliver the target state.

6.5 Discriminant Validity

Structural variables should overlap with established measures but must add held-out prediction beyond beliefs, preferences, switching costs, responsibility, reference points, and generic inertia.

6.6 Information Registration and Belief Updating Must Be Measured Separately

Critical tasks should separately measure comprehension, belief updating, and choice, distinguishing failed information processing from belief revision without action revision and from joint attenuation.

6.7 Prospective Prediction

Baseline structural estimates should precede occupancy assignment and predict later belief and action sensitivity. Primary predictive comparisons should hold out entire participants rather than trials whenever repeated observations are used.

7. A Discriminating Test of Choice Reversal

A decisive experiment creates an incumbent choice, measures beliefs and baseline reversibility, randomizes active occupancy without new task-relevant information, and then crosses structural support with diagnostic evidence. An incentivized investment or project-continuation task is one implementation because beliefs, payoffs, switching fees, and alternatives can be made explicit.

The design must cross initially low and relatively high revisability; a sample consisting only of resistant choices cannot identify the gate. Occupancy and later support are randomized within the pre-occupancy initial-state strata. Figure 4 shows the design and its predicted response displacement.

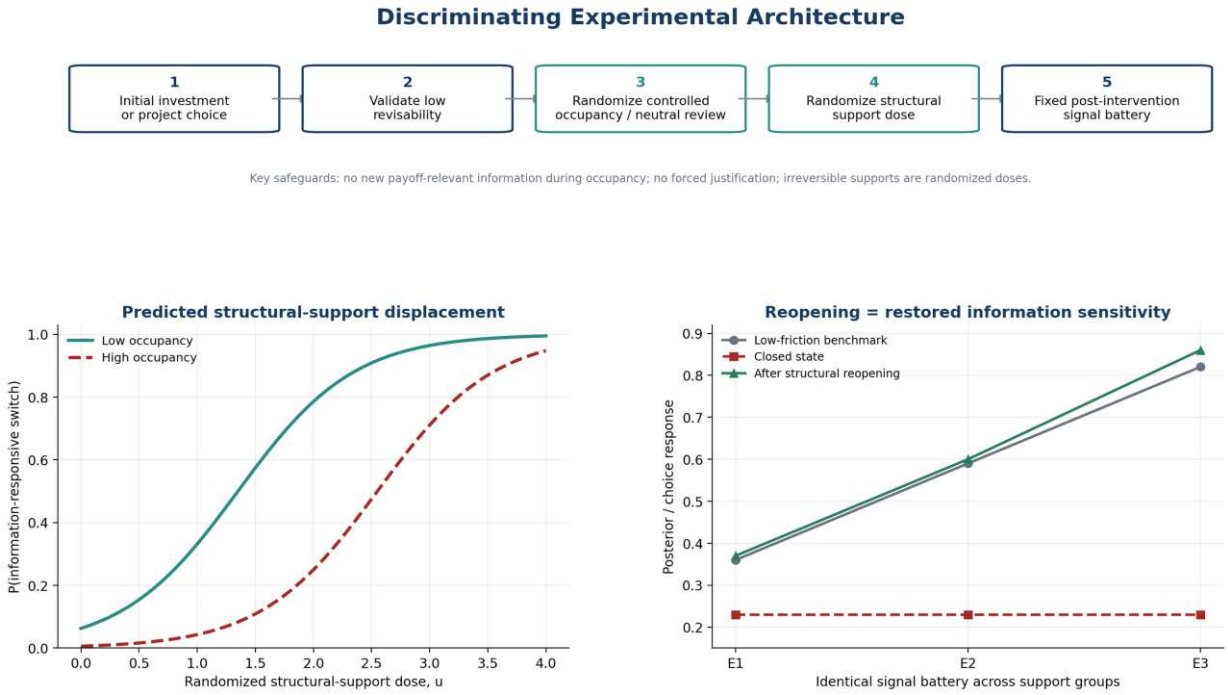


Figure 4. Proposed choice-reversal test. Initial reversibility is assessed before occupancy and support assignment. The same evidence schedule is used across support conditions. The full design crosses low and higher initial revisability; reopening is assessed through information sensitivity.

7.1 Task Calibration and Benchmarks

Pilot work should calibrate information comprehension, baseline commitment, initial-revisability measurement, and support levels so that neither choice nor reversal sits at a ceiling or floor. The confirmatory task should use known priors, signal likelihoods, payoffs, and transition rules.

7.2 Normative and Subjective Benchmarks

For an economic implementation, a Bayesian or expected-value benchmark can be computed from announced priors, signal likelihoods, terminal outcomes, and switching costs. A second benchmark based on elicited beliefs separates belief updating from action selection. Nonlinear utility or option value must be modeled when they are part of the task.

7.3 Initial-State Validation Before Randomization

Initial-state classification must be completed before occupancy is randomized and cannot be based on non-switching alone. It should combine comprehension, current beliefs, documented constraints, and a preregistered response measure calibrated against a low-friction condition.

7.4 Occupancy Without New Evidence or Forced Justification

Neutral choice-recognition cues can reactivate the incumbent while matched control tasks hold duration, attention, and motor response constant. Cues should contain no new outcomes, prices, arguments, or evaluative feedback.

7.5 Randomized Support and Fixed Evidence

After occupancy, assign one prespecified support level to each independent target and independently assign diagnostic evidence. The estimand is a shift in the support-by-evidence response surface, not a simple treatment-induced increase in switching.

7.6 Rival Processes and Confirmatory Comparison

Measure attention, memory, affect, demand awareness, updated reference points, and beliefs as process outcomes. The primary randomized contrast should not condition on post-treatment mediators. Generic linear and saturating activation, reference adaptation, dynamic switching costs, and the restricted commitment gate should be frozen before confirmatory testing.

7.7 Confirmatory Predictions

H1: baseline structural indicators improve held-out prediction of information-sensitive action beyond explicit payoff, belief, preference, and switching-cost variables.

H2: randomized active occupancy shifts the support-by-information response more strongly in initially low-revisability than higher-revisability states, and the interaction improves prediction against an equal-complexity generic saturating rival.

H3: independently validated burden, tolerance, and path measures respond differentially to mechanism-matched interventions after ordinary payoff consequences are represented.

H4: the restricted occupancy model predicts held-out participants' belief and action responses better than static, time-only, generic-saturation, reference-adaptation, and dynamic switching-cost rivals.

7.8 Power, Preregistration, and Falsification

Participant sample size must come from pilot-informed binary or allocation outcomes, classification reliability, treatment compliance, and the smallest substantively relevant interaction. The synthetic recovery exercises in Section 9 are identification analyses, not participant power calculations.

Preregister initial-state criteria, support and evidence ranges, exclusion rules, the primary interaction, multiplicity policy, and predictive model comparison before recruitment. A null randomized occupancy interaction directly limits SRT-2 in the tested domain; weak model discrimination is inconclusive rather than confirmatory.

7.9 Estimands

Let A denote randomized high versus low occupancy, L_0 the pre-occupancy low-revisability stratum, u support, and e diagnostic evidence. Estimate the total occupation effect on $m(A,L_0,u,e) = E[Y|A,L_0,u,e]$ by intention-to-treat. The gate contrast is the difference between occupancy effects on the support-by-evidence response across initial-state strata.

For binary switching, report probability-scale contrasts at actual support and evidence levels; for bounded allocation, use an appropriate response model. Freeze rival models and evaluate held-out predictive scores at the participant level.

8. Implications for Judgment and Decision Making

The framework is intended as a restricted account of history-dependent choice reversal, not a new label for inertia. Its useful question is whether current belief and action are sufficient statistics for future reversibility.

8.1 Beliefs, Preferences, and Transition Constraints

Belief updating, reference dependence, motivated cognition, and status-quo preference remain primary explanations of persistence. SRT adds value only if independently measured transition variables and prior occupancy improve prediction after those state variables are represented.

8.2 Choice Architecture and Mechanism Matching

The distinction between information and transition structure suggests a design principle. Better evidence targets informational failures; fee or consequence relief targets burdens on an existing route; reversibility can alter tolerance; and additional options can create a route. These are competing causal operations, not interchangeable forms of persuasion.

9. Computational Identification and Design Requirements

The computational analyses ask whether the proposed commitment gate is identifiable under candidate experimental designs. They use synthetic data only and do not count as evidence that the mechanism describes human decision making.

Generative-AI transparency. OpenAI ChatGPT and Anthropic Claude were used iteratively for drafting and editorial revision, literature-search planning, and assistance with computational code development and checking (Anthropic, 2026; OpenAI, 2026). All cited sources and reported computational outputs were reviewed and the supplied scripts were rerun. The author is responsible for the final theory, code, interpretation, and manuscript.

9.1 Preliminary Recovery Diagnostics

Earlier synthetic checks compared static, time-only, generic activation, and occupancy-saturation generators and examined parameter recovery under favorable direct measurement. Those analyses establish only that the specified response shapes can sometimes be distinguished. Because the generic alternative was not equally flexible, they do not identify the commitment gate. Full specifications and outputs remain in the supplementary archive.

9.2 Exact Identification Limit of Fixed Accessibility

The reduced gated curve is $U = \alpha + A[1-\exp\{-\rho dq(1-K_0)\}]$. At constant K_0, define $\rho' = \rho(1-K_0)$. The generic curve $U = \alpha + A[1-\exp(-\rho' dq)]$ then gives identical predictions for every d and q. No increase in N resolves this equivalence. The supplied reproducibility archive includes an exact-equivalence self-test. A discriminating test therefore needs independently assessed variation in initial accessibility and a generic rival allowed the same saturation form and comparable flexibility. Crossing initial-state conditions provides such variation in principle; whether the available range is sufficient remains an empirical design question.

9.3 Binary-Choice Comparison With an Equal-Complexity Saturation Rival

A diagnostic simulation observes binary switching rather than a directly observed reopening dose. Both candidates have seven fitted parameters and the same support and evidence response. For initial accessibility k, $U = a + b_k(k-.12) + A[1-\exp(-\rho h)]$, where $h = dq$ for generic saturation and $h = dq(1-k)$ for the commitment gate.

$P(\text{switch}) = \text{logistic}\{c + b_u u + b_e e \cdot \text{logistic}[2(u-U)]\}$.

The direct support term allows support to increase switching without restoring evidence sensitivity; the inner evidence link changes sensitivity to signed diagnostic evidence, and the scale 2 is fixed in both models.

The base design crosses d = {0,.5,1.5,3}, q = {.15,.85}, k = {.12,.70}, u = {0,1,2,3}, and e = {−1,0,1}. Each of 192 cells contributes 16 independent binary events split equally between fitting and held-out scoring. There are 80 datasets per generating model.

Correct selection is 51/80 (63.8%; Wilson 95% CI 52.8–73.4%) for generic saturation and 42/80 (52.5%; 41.7–63.1%) for the gated generator. Thus, merely crossing two accessibility levels does not make this narrow design reliably discriminating. Pooled over both generators, base correct selection is 93/160 (58.1%; 50.4–65.5%); the larger 800-dataset rerun of the same base design in Section 9.4 gives 63.8%. Table 2 reports both generators under the base and the exploratory wider design.

Accessibility is known without error, events are independent, the fitted family is correctly specified, and held-out events share training design cells. The exercise is therefore a design diagnostic, not a power estimate or behavioral result.

Table 2. Synthetic binary choice model recovery under equal-complexity saturation rivals.

Design	Generating model	Correct selection	Wilson 95% interval
Base	Generic saturation	51/80 (63.8%)	52.8–73.4%
Base	Gated saturation	42/80 (52.5%)	41.7–63.1%
Wider exploratory	Generic saturation	74/80 (92.5%)	84.6–96.5%

Design	Generating model	Correct selection	Wilson 95% interval
Wider exploratory	Gated saturation	74/80 (92.5%)	84.6–96.5%

9.4 Which Design Dimension Drives Recovery?

The same generator, likelihood, and fitting code were rerun while widening one design dimension at a time and holding the other three at base values. Cell count, events per cell, generating parameters, and candidate families remain fixed.

Table 3 reports the results. Base and all-widened designs yield 63.8% and 93.0% pooled correct selection over 800 datasets each; the all-widened row re-estimates the exploratory wider design of Table 2 (148/160, 92.5%) with five times as many datasets and an independent seed. Widening initial accessibility from {.12,.70} to {.05,.95} alone raises recovery to 83.9%, roughly two-thirds of the total gain. Widening maximum dwell from 3 to 6 adds 0.3 percentage points; widening evidence adds 6.9; widening support lowers recovery by 4.1 because the zero-support anchor is lost.

This pattern follows the analytic equivalence: the gate multiplies the saturation rate by (1−k), whereas dwell and activation enter both models identically. Longer commitment exposure therefore does not discriminate the gate from this rival. Stronger, independently measured variation in initial reversibility does. The 83.9% figure is specific to this comparison and does not establish unique gate identification. Crossing initial accessibility is necessary but not sufficient: a sufficiently flexible rival in which k also modifies the saturation rate could still encompass the gated restriction.

Table 3. One-factor-at-a-time decomposition of the design improvement relative to the base binary-choice comparison in Section 9.3. Each row widens one design dimension to its exploratory range and holds the remaining three at base values. Correct selection is pooled across both generating models with 400 datasets each; intervals are Wilson 95% intervals on the pooled proportion. Cell count, event count, generating parameters and candidate families are identical across rows. Synthetic data; seeds 20261011–20261016; no optimizer failures.

Design	Generic	Gated	Pooled (Wilson 95%)	Δ vs base
Base (anchor)	264/400	246/400	63.8% (60.4–67.0)	—
Dwell d: max 3 → 6	267/400	245/400	64.0% (60.6–67.3)	+0.3
Support u: {0–3} → {.5–3.5}	238/400	239/400	59.6% (56.2–63.0)	−4.1
Evidence e: ±1 → ±2	290/400	275/400	70.6% (67.4–73.7)	+6.9
Accessibility k: {.12,.70} → {.05,.95}	350/400	321/400	83.9% (81.2–86.3)	+20.1
All four widened	384/400	360/400	93.0% (91.0–94.6)	+29.3

10. Discussion

The completed contribution is an identification result. Fixed initial accessibility makes a commitment gate algebraically indistinguishable from generic saturation at any sample size; crossing two accessibility levels remains weak in the narrow binary design; and most of the recoverability gained by widening the design comes from the sampled accessibility range. The behavioral proposition—that commitment history can alter future transition requirements—remains untested.

For decision research, the practical implication is direct: a study cannot identify history-dependent gating merely by randomizing how long or how often a choice is reactivated while holding starting reversibility fixed. The design must deliberately vary or sample initial reversibility and compare the gate with equally flexible state-history rivals.

10.1 What the Theory Adds

The formal reachability condition is inherited from control and viability theory. The proposed decision contribution lies in a target-specific decomposition of transition burden, adjustment tolerance, and path availability, together with the hypothesis that active low-revisability occupancy can modify those constraints.

The current landscape determines accessibility; occupancy may alter the later landscape. SRT-2 therefore requires a randomized response shift plus predictive discrimination against generic saturation and other state-history models. Mechanism claims additionally require independent measurement of the proposed structural variables.

10.2 What the Identification Result Changes

The identification result narrows the empirical strategy. Fixed-accessibility dwell curves are not tests of the gate, and favorable fit to a saturating curve is not evidence for structural plasticity. Initial-state variation must enter the design as a measured or manipulated dimension rather than a nuisance variable held constant.

The simulations then show that weak variation can still be insufficient. In the present comparison, widening accessibility dominates longer dwell as a source of recoverability. Pilot work should therefore establish whether the available initial-state range is broad and reliable enough before a confirmatory test is launched.

10.3 Choice State Is Not Necessarily Future Trajectory

Crossing a low-revisability boundary does not imply lock-in. The structural margin can recover, remain metastable, or deepen; timing and reversibility remain empirical questions.

A person can also update locally while one discrete transition remains difficult. SRT is a directed-transition model, not a trait theory of irrationality or global inflexibility.

10.4 Implications for Choice Architecture

If the mechanism is supported, the framework would distinguish interventions that improve information processing from interventions that alter implementation conditions. The relevant outcome is appropriate information sensitivity, not maximal switching.

10.5 Rational Persistence and Welfare

Structural persistence is not inherently biased. Transaction costs, risk, coordination requirements, or expected losses can make staying optimal. Any welfare claim requires an explicit normative benchmark.

10.6 Constraints on Generality

The present formal and computational results concern a restricted class of directed choice-reversal models and a specified generic saturation rival. The same logic can apply to economic, consumer, organizational, or policy decisions, but no human domain is validated by the synthetic analyses. The relative importance of initial accessibility may change under different observation models, noisy measurement, clustered participants, learning, or richer state-history rivals.

10.7 Final Empirical Criterion

The empirical question is whether occupying a constrained choice state changes later information-sensitive reversal requirements beyond established valuation, learning, and state-history processes.

Evidence requires reproducible initial-state moderation, fair held-out prediction against equal-complexity rivals, and independent support for any burden-tolerance-path mechanism. Otherwise a narrower activation or dynamic switching-cost account should be preferred.

The history of constrained commitment may change the support required for later information-responsive choice reversal.

The article proves which design class cannot identify that proposition and identifies the sampled dimension that most improves recovery against the specified rival.

Transparency and Open Practices

All reported computations use synthetic data; no human observations are analyzed. An anonymized reproducibility archive is supplied with the masked review materials and contains the recovery code and outputs, the gate-equivalence self-test, binary-choice analyses, one-factor-at-a-time decomposition, deterministic seeds, and checksums. Public repository details are withheld during masked review and will be restored in the final article. An earlier version of the archive, covering the preliminary diagnostics summarized in Section 9.1, is publicly deposited; its identifier is withheld during masked review.

Generative AI use is disclosed at the beginning of Section 9 and cited in the reference list. The AI tools assisted with iterative drafting, editorial revision, literature-search planning, and code development/checking; responsibility for verification and interpretation remains with the author.

Ethics. The reported work is theoretical and computational and uses only synthetic data. No human participants or animals were studied. Any future behavioral experiment would require prospective ethics review and informed consent.

References

Akerlof, G. A., & Dickens, W. T. (1982). The economic consequences of cognitive dissonance. American Economic Review, 72(3), 307-319.

Ambuehl, S., & Li, S. (2018). Belief updating and the demand for information. Games and Economic Behavior, 109, 21-39. https://doi.org/10.1016/j.geb.2017.11.009

Anthropic. (2026). Claude [Large language model]. https://claude.ai/

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. Organizational Behavior and Human Decision Processes, 35(1), 124-140. https://doi.org/10.1016/0749-5978(85)90049-4

Aubin, J.-P. (1991). Viability theory. Birkhäuser.

Aubin, J.-P. (2001). Viability kernels and capture basins of sets under differential inclusions. SIAM Journal on Control and Optimization, 40(3), 853-881. https://doi.org/10.1137/S036301290036968X

Bachmann, K. (2024). Do you have a choice?: Implications for belief updating and the disposition effect. Journal of Economic Psychology, 102, 102718. https://doi.org/10.1016/j.joep.2024.102718

Barberis, N., & Xiong, W. (2012). Realization utility. Journal of Financial Economics, 104(2), 251–271. https://doi.org/10.1016/j.jfineco.2011.10.005

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2019). A network perspective on attitude strength: Testing the connectivity hypothesis. Social Psychological and Personality Science, 10(6), 746-756. https://doi.org/10.1177/1948550618781062

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. Psychological Review, 123(1), 2-22. https://doi.org/10.1037/a0039802

Drobner, C., & Goerg, S. J. (2024). Motivated belief updating and rationalization of information. Management Science, 70(7), 4583-4592. https://doi.org/10.1287/mnsc.2023.02537

Friedman, D., & Wang, Z. (2026). Motivated beliefs meet in the marketplace. Journal of Behavioral and Experimental Economics, 123, 102578. https://doi.org/10.1016/j.socec.2026.102578

Frydman, C., & Rangel, A. (2014). Debiasing the disposition effect by reducing the saliency of information about a stock's purchase price. Journal of Economic Behavior & Organization, 107, 541–552. https://doi.org/10.1016/j.jebo.2014.01.017

Grasman, R. P. P. P., van der Maas, H. L. J., & Wagenmakers, E.-J. (2009). Fitting the cusp catastrophe in R: A cusp package primer. Journal of Statistical Software, 32(8), 1-27. https://doi.org/10.18637/jss.v032.i08

Guney, B., & Richter, M. (2018). Costly switching from a status quo. Journal of Economic Behavior & Organization, 156, 55-70. https://doi.org/10.1016/j.jebo.2018.08.017

Henry, T. R., Robinaugh, D. J., & Fried, E. I. (2022). On the control of psychological networks. Psychometrika, 87(1), 188-213. https://doi.org/10.1007/s11336-021-09796-9

Odean, T. (1998). Are investors reluctant to realize their losses? Journal of Finance, 53(5), 1775–1798. https://doi.org/10.1111/0022-1082.00072

OpenAI. (2026). ChatGPT [Large language model]. https://chatgpt.com/

Quispe-Torreblanca, E., Gathergood, J., Loewenstein, G., & Stewart, N. (2025). Investor logins and the disposition effect. Management Science, 71(1), 219–239. https://doi.org/10.1287/mnsc.2022.00359

Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. Quarterly Journal of Economics, 114(1), 37-82. https://doi.org/10.1162/003355399555945

Robalo, P., & Sayag, R. (2018). Paying is believing: The effect of costly information on Bayesian updating. Journal of Economic Behavior & Organization, 156, 114-125. https://doi.org/10.1016/j.jebo.2018.09.016

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. Journal of Risk and Uncertainty, 1, 7-59. https://doi.org/10.1007/BF00055564

Stocker, J. E., Koppe, G., Reich, H., Heshmati, S., Kittel-Schneider, S., Hofmann, S. G., Hahn, T., van der Maas, H. L. J., Waldorp, L., & Jamalabadi, H. (2023). Formalizing psychological interventions through network control theory. Scientific Reports, 13, 13830. https://doi.org/10.1038/s41598-023-40648-x

Talpsepp, T., & Vaarmets, T. (2019). The disposition effect, performance, stop loss orders and education. Journal of Behavioral and Experimental Finance, 24, 100240. https://doi.org/10.1016/j.jbef.2019.100240

van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden transitions in attitudes. Sociological Methods & Research, 32(2), 125-152. https://doi.org/10.1177/0049124103253773

Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. Journal of Economic Behavior & Organization, 33(2), 167–184. https://doi.org/10.1016/S0167-2681(97)00089-9

Zimmermann, F. (2020). The dynamics of motivated beliefs. American Economic Review, 110(2), 337-361. https://doi.org/10.1257/aer.20180728

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DecisionSRTBlindedSupplement.docx
- DecisionSRTReproducibilityFORREVIEW.zip