Two Timescales, Two Calibers: Auditing Consumer Adaptation in Algorithmically Mediated Environments

Supplementary Materials

Manuscript information-4549232 (R2) — Supplementary Materials.

Numbering: supplementary tables, figures, sections, and algorithms are numbered S1, S2, … independently of the main text. Items labeled "archived" or "legacy" document the retired split-half provenance analysis and the scale-level calibration of Equation (1); they are retained for provenance only and do not define the revised matched-estimand analysis.

Relationship to code and data: analysis code, frozen outputs, and provenance documentation are archived in the project code repository (GitHub/Zenodo record cited in the main text). JDsearch and JData raw records are governed by data-use agreements and cannot be publicly redistributed; the schemas, screening chains, and reconstruction rules below permit audit of the analytical samples without access to the raw records.

Contents: S1. Data Provenance and Variable Reconstruction (Tables S1–S8); S2. Computational Verification and Generative Implementation (Sections S2.1–S2.8, Tables S9–S11, Algorithm S1, Figures S1–S4); S3. Full Portability, Measurement, and Auxiliary Audits (Tables S12–S20, Figures S5–S7).

S1. Data Provenance and Variable Reconstruction

Section S1 documents the raw data structure, sample-screening pipelines, and feature-engineering chains of the three empirical studies. It provides reconstruction details and supporting descriptive quantities for analyses reported in the main text; these materials do not introduce additional substantive hypothesis tests or alter the evidentiary hierarchy of the paper. All reported quantities are reproduced from the frozen analysis package without modification.

S1.1. Study 1 (JDsearch)

Table S1. Schema of the raw data files used in Study 1 (JDsearch).

Raw file	Field	Description	Role in analysis
user_behavior_data.txt	queries	Search-query identifiers (underscore-separated sequence per user row)	Behavioral content
	wids	Product identifiers (underscore-separated sequence)	Product-level resolution via product_meta_data
	types	Behavior-type codes (underscore-separated sequence)	Click/order classification
	times	Event timestamps (underscore-separated sequence)	Within-user temporal ordering for archived half-split and revised wave constructions.
product_meta_data.txt	wid	Product identifier	Join key to behavior rows
	brand	Brand identifier	Brand-level aggregation (maxshare, HHI)
	cate1	Category level-1 identifier	Category-level Shannon entropy

Raw file	Field	Description	Role in analysis
user_features.csv	click_count	Cumulative click volume	Archived provenance branch only; not used in revised REWB CDI.
	shannon_entropy	Cross-sectional category entropy	Archived provenance branch only; not used in revised REWB CDI.
	n_unique_cates	Number of unique categories	Archived cross-sectional validity screen only.

Note. user_behavior_data.txt is 957 MB (tab-delimited); product_meta_data.txt is 4.84 GB. Both files were obtained under a data-use agreement. Only fields used in the present analysis are listed. The `times` field supplies within-user event ordering. In the archived provenance analysis, each ordered sequence was divided into first and second halves. In the revised primary panel, the ordered sequence is partitioned into observable within-user waves as described in Section 4.1. These wave indices represent relative sequence position rather than reconstructed calendar time.

Table S2. Screening pipeline of the archived split-half Study 1 provenance sample ($N = 25{,}467$).

Step	Rule	Users
A1	Valid parsing (len(p) ≥ 7)	173,831
A2	Behaviors per user n ≥ 10	157,417
A3	≥ 3 clicks in each half of the observation span (entropy computable in both halves)	39,960
A4	≥ 3 first-half orders (consolidation proxy computable); analytic sample	25,467
X1	Cross-sectional branch: users with ≥ 1 category-level behavior (n_unique_cates > 0)	158,085

Note. This table documents the retired split-half provenance analysis only and does not define the sample construction of the revised primary REWB panel. Each user's behavior sequence is split at its midpoint (mid = n//2) into first and second halves. No 365-day observation-window filter was applied. The product catalog covers 88.80% of demanded product identifiers (3,131,243 of 3,525,977 wid); unmatched identifiers are skipped in entropy and uratio computations. Per-user observation windows extend up to 635 days (20.9 months; mean = 235 days, median = 286 days).

Table S3. Feature engineering and variable construction for Study 1 (primary REWB panel).

Analysis variable	Raw input	Aggregation key	Construction logic	Role
$maxshare_{it}$	wids, brand	user × wave	Maximum brand click share within each observable wave (Eq. 2)	IV - consolidation (primary representation)
HHI_{it}	wids, brand	user × wave	Herfindahl–Hirschman index of within-wave brand click shares (Eq. 3)	IV - consolidation (representation audit)
H_{it} (entropy)	wids, cate1	user × wave	Shannon entropy of within-wave category click distribution (Eq. 4)	DV - behavioral diversity
log(click volume)$_{it}$	wids (click type)	user × wave	Log of within-wave valid behavioral-event count (pipeline variable n_click)	Covariate / scale control

Note. The archived split-half variable constructions (Δentropy, Δuratio) of the retired provenance analysis are documented in Tables S12–S13; the present table lists only the wave-based construction of the primary REWB panel.

Note. All primary-panel measures are constructed within observable waves using catalog-matched clicks. Maxshare and HHI are computed from within-wave brand shares, whereas Hit is Shannon

entropy over within-wave category shares. Archived split-half variables (Δentropy, Δuratio) belong only to the retired provenance analysis and are documented separately in Tables S12–S13.

S1.2. Study 2 (Amazon Reviews)

Table S4. Schema of the raw data file used in Study 2 (Amazon Reviews 2023, Electronics).

Field	Description	Role in analysis
rating	Review star rating (integer 1–5)	DV
timestamp	Review timestamp (Unix)	Tenure clock (years since first observed review); year filter 1996–2023
parent_asin	Parent product identifier	Validity filter (non-missing); product × year demeaning
asin	Product identifier	Product-level demeaning (legacy track); product × year cells
user_id	User identifier	User clustering; user-level aggregation; era assignment (first-review year)
verified	Verified-purchase flag	Robustness check R1 (verified-only subsample)

Note. The raw file Electronics.jsonl.gz is 6.03 GB and spans 1996-11-18 to 2023-09-13 (approximately 27 years). Only fields used in the present analysis are listed.

Table S5. Variable construction: rating tracks, tenure, and era for Study 2.

Variable	Construction	Role
y_raw	Raw review star rating	Track 1 - raw ratings
y_itemyear	Rating demeaned within product × calendar year (primary standardized track)	Track 2 - primary B track
y_year	Rating demeaned within calendar year	Track 3 - C track
y_item	Rating demeaned within product (full-period mean; legacy track)	Robustness track (demoted from primary)
tenure	Years since the first observed review	Time variable in the quadratic growth model
era	Cohort by first-review year: EC1 1996–2003; EC2 2004–2017; EC3 2018–2023	Era decomposition

Note. The primary B-track baselines are computed from the full S1 valid sample: 43,886,942 reviews, 3,839,795 product × year cells, and a merge coverage of 1.000000; cell review counts are p10 = 4, p50 = 69, p99 = 25,554, and among analysis reviews, 4.37% fall in single-review cells and 88.12% in cells with n ≥ 5. Calendar-year attribution is used for the item-year demeaning. Consistent with Section 5.2.

S1.3. Study 3 (JData)

Table S6. Schema of the raw data files used in Study 3 (JData).

Raw file	Field	Description	Role in analysis
JData_Action_201602/03/04.csv	user_id	User identifier	User-level aggregation
	sku_id	Item (SKU) identifier	Unique-item counts (n_unique_items)
	time	Event timestamp	Monthly scan (2016-02 to 2016-04)
	type	Behavior type (1 browse by page; 2 browse by product; 3 add to cart; 4 cart removal; 5 purchase; 6 favorite)	Active-user construction (≥1 type-1 and ≥1 type-5); retained types {1,2,4,5,6}
	cate	Category identifier	Category diversity (n_unique_cates)
	brand	Brand identifier	Brand diversity (n_unique_brands)
User_table.csv	user_id	User identifier	Join key (left-merge)

Raw file	Field	Description	Role in analysis
	age	Age bracket (1–6: 16–25, 26–35, 36–45, 46–55, 56–65, 65+; −1/NaN = missing)	Validity screen ($age > 0$)
	sex	Sex (0 male; 1 female; 2 undisclosed; NaN = missing)	Gender comparison subset (sex ∈ {0, 1})
	user_lv_cd	User level code	Consumption-level analysis

Note. The three action tables total approximately 2.3 GB. The summary file jdata_user_summary.csv (26,244 users × 9 behavior columns: n_actions, n_view, n_cart, n_fav, n_order, n_follow, n_unique_items, n_unique_cates, n_unique_brands) was constructed from the action tables and left-merged with the user table (105,321 rows; 100% match). Consistent with Section 6.1.

Table S7. Sample construction and attribute validity pipeline for Study 3.

Step	Rule	Users
S0	Users with ≥ 1 type-1 and ≥ 1 type-5 behavior (seed-42 sampling cap of 50,000 not triggered)	26,244
S1	diversity_score = log(1 + n_unique_cates) + log(1 + n_unique_brands) computed over retained types {1,2,4,5,6}	26,244
S3	Validity screen:$age > 0$ (effective sample; the sex non-missing clause is vacuous)	23,644
S4	Gender comparison subset: sex ∈ {0, 1} (male/female) within the effective sample; sex = 2 undisclosed excluded	13,081

Note. Step numbering follows the frozen analysis pipeline, in which S2 is an intermediate computation stage rather than a screening step. Consistent with Section 6.1.

The original screening log distinguishes intermediate computation stages from exclusion steps. Among 26,244 active users, 2,600 fail the age-validity screen (2,599 with age coded −1 and 1 with missing age), leaving 23,644; the single user with missing sex falls within the removed set, so the totals reconcile exactly. The age distribution of the screened users is {−1: 2,599; 1: 3; 2: 2,180; 3: 12,763; 4: 7,434; 5: 802; 6: 462; missing: 1}, and monthly scans of the action tables record views of 72,006 / 100,819 / 105,172 and type-5 behaviors of 9,612 / 21,947 / 26,244 across February–April 2016.

The sex-code audit identifies 0 as male, 1 as female, and 2 as undisclosed. The full sex distribution under the corrected codebook is {0 (male): 11,750; 1 (female): 2,334; 2 (undisclosed): 12,159; missing: 1}; the valid male–female comparison includes 10,945 and 2,136 users, respectively, within the effective sample (sex = 2 undisclosed, n = 10,563, excluded). The retired contrast based on codes 1 and 2 is preserved only to document the corrected provenance.

Table S8. Feature engineering and variable construction for Study 3.

Analysis variable	Raw input	Aggregation key	Construction logic	Role
diversity_score	n_unique_cates, n_unique_brands	user	log(1 + n_unique_cates) + log(1 + n_unique_brands)	DV - behavioral breadth composite (log category + log brand counts)
consumption level	User_table.user_lv_cd	user	User-level code	IV - consumption level
age	User_table.age	user	Age bracket (1–6)	Individual-difference covariate
sex	User_table.sex	user	0 male; 1 female; 2 undisclosed	Individual-difference covariate
high/low tier	diversity_score	user	Low ≤ median < high (median = 4.2767)	Stratified contrast

Note. Unique category and brand counts are computed over the full retained-type set {1, 2, 4, 5, 6}. The stratified high/low contrast is defined on the effective sample ($N = 23{,}644$) with the disclosed median split; the semantic interpretation of the n_order column is subject to the type-mapping caveat of the data package and is avoided in interpretation. Consistent with Section 6.2.

S2. Computational Verification and Generative Implementation

This section documents the disclosed decision rules, estimation configuration, fidelity gates, and verification procedures of the three studies. No new numerical results are introduced beyond those reported in the main text.

S2.1. Disclosed Decision Rules

Disclosed decision rules of the three studies (full record).

Study	Rule	Consequence
Study 1 (archived)	If the full-sample ρ has the same sign as the pilot and $\lvert\rho\rvert \in$ [0.08, 0.20]	The full-sample value becomes the estimate of record; pilot values are downgraded to pre-validation (this branch was triggered)
Study 1 (archived)	If $\lvert\rho\rvert < 0.08$ but significant, same sign	Dual-report and discuss effect-size attenuation (large-sample dilution)
Study 1 (archived)	If sign reversal or non-significance	The study is downgraded and reported as-is
2	The inverted-U hypothesis is supported iff both standardized tracks (y_itemyear and y_year) are inverted-U (Bonferroni$\alpha = .025$; turning point in range)	Not supported: both tracks are U-shaped
3	All four core tests reproduced (same sign; magnitude deviation < 20%); zero-anchor rule: |re-run d| < 0.02 counts as reproduction	Frozen values retained with zero change

Note. The Study 1 branch rules document the retired split-half provenance analysis and did not govern selection, estimation, or interpretation of the revised REWB CDI.

S2.2. Study 2: Dual-Path Estimation Disclosure

Study 2 separates computational verification from estimand sensitivity. Path A is the designated review-level analysis, and Path A′ independently reimplements that estimator for computational verification. Path B re-estimates the same trajectory form on equally weighted user × tenure-year cells and therefore targets a different population quantity

rather than serving as a computational replication. The choice of weighting scheme determines which population quantity is being estimated [70,71], and the two need not agree in sign: under Path B, the pooled raw and year-demeaned tracks exhibit statistically significant negative curvature, the pooled item-demeaned track (the auxiliary product-level-demeaned outcome) has positive curvature that does not meet the prespecified Bonferroni-adjusted significance threshold ($\alpha = .025$), and the primary item × year-demeaned track yields a small negative coefficient ($\pi_2 = -6.21 \times 10^{-5}$, p = .0417) that likewise does not meet the threshold. Full numerical concordance across all checks was therefore not attained; this is an estimand difference, not a computational error, and the corresponding estimates are reported in Section S2.3. Per the disclosed protocol, the review-level specification is primary, both paths are reported, and because user-weighted re-estimation flips the sign of the curvature, no directional claim is made about the population-level trajectory.

S2.3. Study 2: Estimation Configuration, Coverage, and Verification

The primary review-level model is estimated by OLS with user-clustered robust standard errors (statsmodels 0.14.6), and an independent numpy reimplementation (Path A′) verifies it to machine precision, with a maximum absolute difference of 6.2×10^{-14} across all reported coefficients. On the primary item × year-demeaned track, Path A yields $\pi_2 = +1.157 \times 10^{-4}$ ($p = 1.9 \times 10^{-4}$), whereas the equally weighted Path B yields $\pi_2 = -6.21 \times 10^{-5}$ ($p = .0417$); across the 12 comparisons, 9 agree in the sign of π_2 and 7 in statistical significance (all_pass = False). The 12 comparisons comprise the four scopes (pooled and the three first-review eras) crossed with the three historical outcome tracks (raw, item-demeaned, and year-demeaned); the pooled Path A/B comparison on the primary item × year-demeaned track is reported separately above and is not counted among the 12. The pooled Path B results are sensitive to the outcome construction: the raw and year-demeaned tracks have statistically significant negative curvature, whereas the auxiliary item-demeaned track has positive curvature that, like the primary track's negative coefficient reported above, does not meet the Bonferroni-adjusted threshold ($\alpha = .025$). Three robustness layers bracket ceiling and heterogeneity concerns. The first is an ordered logistic regression on a 200,000-review subsample, which tests robustness to ceiling effects. The second is a two-stage random-coefficient analysis that fits per-user quadratic trajectories for the 904,403 users with at least four cells. The third is an auxiliary SEM layer (semopy 2.3.11; 20,000 users × tenure 0–9-year cells). It serves only as a variance-structure diagnostic, because it does not return the latent-mean structure required for the focal trajectory question.

S2.4. Study 3: Verification and Data-Level Reconstruction

The statistical Path A′ reimplementation (hand-written numpy, same estimator) passed all ten checks (10/10), with a maximum difference of 2.16×10^{-12}. At the data level, the 26,244-user, nine-column summary was fully rebuilt from the original three action tables (JData_Action_201602/03/04.csv, approximately 2.3 GB in total); all nine columns matched exactly (26,244/26,244 rows; maximum absolute difference 0.0). Attribute provenance was verified field by field: age, sex, and user_lv_cd match the user table (105,321 rows) at 100%, and recomputing active users (at least one type-1 and one type-5 behavior) reproduces 26,244, confirming that the seed-42 sampling cap was never triggered.

A timestamp provenance audit of 50,601,736 parsed action rows (0 unparseable) confirms that event timestamps span 2016-02-01 to 2016-04-15 (approximately ten weeks), with per-file ranges of 01-31–02-29 (201602), 02-29–03-31 (201603), and 03-31–04-15 (201604), plus a single pre-window boundary record at 2016-01-31 23:59:02.

The sex-codebook correction is recorded here in full: the corrected reading is 0 = male, 1 = female, and 2 = undisclosed; the legacy codebook (1 = male, 2 = female, 0 = missing) and the legacy sex $\in \{1, 2\}$ gender test are retired, and the corrected male–female comparison is reported in Section 6.3. Analyses based on age, behavioral opportunity, behavioral breadth, and other variables not conditioned on sex are unaffected by this correction. Finally, because p-values with $N > 20{,}000$ fall below the double-precision floor, CSV values of 0.0 denote $p < 1 \times 10^{-300}$.

Table S9. Mathematical notation and variable definitions.

Symbol	Definition	Equation / location	Study
$x_t,\ x_0$	Expressed preference breadth at period t; fixed-point breadth	Eq. (1), §2.3	Theory
a_t	Accuracy of the recommendation stream at period t	Eq. (1), §2.3	Theory
$\lambda,\ \delta$	Fast-channel gain scaling; slow-channel consolidation rate	Eq. (1), §2.3	Theory
$g(a_t)$	Within-period efficiency gain of accuracy, g ≥ 0	Eq. (1), §2.3	Theory
c_{ibt}	Consumer i's catalog-matched clicks on brand b in wave t	§3.3; Table S3	S1
s_{ibt}	Within-wave share of brand b in consumer i's clicks	§3.3	S1
$maxshare_{it}$	Maximum within-wave brand click share (primary consolidation representation)	Eq. (2), §3.3; Table S3	S1
HHI_{it}	Herfindahl index of within-wave brand click shares (representation audit)	Eq. (3), §3.3; Table S3	S1
H_{it}	Shannon entropy over within-wave category click shares	Eq. (4), §3.3	S1
$\Delta entropy_i$	Archived split-half change measure (retired provenance analysis only; not part of the primary REWB construction)	§4.3; Tables S12–S13	S1
$\Delta uratio_i$	Archived split-half unique-ratio change measure (provenance analysis only)	§4.3; Tables S12–S13	S1
$r_{ijt},\ \tau_{ijt}$	Star rating; user tenure in years since first review	Eq. (14), §5.2	S2
$\pi_{0i},\ \pi_{1i},\ \pi_{2i}$	User-level intercept, slope, curvature; $\pi_{2i} < 0$ = inverted-U	Eq. (14), §5.2; Table 9	S2
EC1 / EC2 / EC3	First-review cohorts by year: 1996–2003 / 2004–2017 / 2018–2023	§5.1; Table 9	S2
$score_i$	Behavioral-breadth composite: log(1 + unique categories) + log(1 + unique brands)	Eq. (15), §6.2; Table S8	S3
$C_i,\ B_i$	Counts of unique categories and brands over retained behavior types	Eq. (15), §6.2	S3
ϱ, partial ϱ	Spearman rank correlation; partial rank correlation controlling log click volume (textbook definitions, not restated)	§4.2–4.3	S1
$\overline{X}_i,\ X_{it}^W$	Consumer-specific mean (between component); wave-specific deviation (within component)	Eqs. (5), (6), §3.4	S1
$\beta_B,\ \beta_W$	Between- and within-consumer associations (joint REWB/Mundlak model)	Eq. (7), §3.5	S1
CDI	Caliber Divergence Index, between-minus-within contrast	Eq. (8), §3.6	S1
CDI^*	Standardized companion, CDI scaled by the predictor/outcome standard-deviation ratio; no universal threshold assigned	Eq. (10), §3.6	S1
$N_{it}^{act},\ A_{it}$	Wave-level valid behavioral-event count; behavioral opportunity	Eq. (11), §3.8	S1
E_{it}	Elapsed months on each consumer's own relative clock	Eq. (13), §3.8	S1
$C_k,\ D(C_k)$	Analytical construction (opportunity, temporal, representation); contrast under that construction	§3.7	S1
$w_{it},\ \theta_i,\ z_i,\ s_i,\ \tau_i$	Generative-probe state, taste, anchor, susceptibility, activity propensity (synthetic only)	Eqs. (18)–(22), §7.2	Probe

S2.5. Simulation Pseudocode

Algorithm S1. *Agent-based world (one replication).*

Initialize: draw activity trait τ_i *~ LogNormal(0, 0.5²), normalized to mean 1; susceptibility* s_i = 0.85 *with probability 0.30, else 0.25; genuine taste* θ_i *~ Dirichlet(0.5·*1_K*); fossil anchor* z_i = 0.8·*(top-3 of* θ_i*) + 0.2/K; consolidation weight* w_i = 0.
For each session t = 1, …, 100:
(1) With probability λ_g*, shock taste:* $\theta_i \leftarrow 0.6\theta_i + 0.4 \cdot Dirichlet(0.5 \cdot 1_K)$.
(2) Expressed profile $x_i = (1 - w_i)\theta_i + w_i\, z_i$*, renormalized.*
(3) Engagement n_i *~ Poisson(*$3\tau_i$*); the realized number of choices can differ from* n_i.
(4) Choices c_i *~ Poisson(*$x_i \cdot n_i$*) (independent Poisson draws; conditional on the realized total, multinomial).*
(5) Consolidate: $w_i \leftarrow min(w_i + \delta \cdot n_i \cdot s_i \cdot (1 + w_i), 0.95)$.
(6) Dynamic-anchor condition only: retrain the anchor $z_i \leftarrow 0.98 \cdot z_i + 0.02 \cdot (c_i / \Sigma c_i)$*; sessions with no realized choices leave the anchor unchanged.*
(7) Accumulate choice counts into first-half ($t \leq 50$*) or second-half (*$t > 50$*) bins.*
After session 100: compute the within-person association ρ between first-half concentration (HHI) and entropy growth, the between-person association r between cumulative engagement and mean entropy, and their cross-sectional operationalization contrast $\Delta_{op} = r - \rho$*. Because the two constituent associations use different behavioral representations and estimands,* Δ_{op} *is retained solely for comparison with the archived simulation target and is not a CDI under the matched-estimand definition in Section 2.2. Ideal condition: step (5) is skipped (*$w \equiv 0$*). Random condition:* x_i = $1_K/K$ *in step (2). The simulation is implemented in Python 3.12.12 (NumPy 2.2.5, PCG64 bit generator via numpy.random.default_rng; SciPy 1.16.2 for rank and product-moment correlations).*

Figure S1 reports the trajectory diagnostic behind the no-warm-up argument of Section 7: consolidation starts from the natural initial condition w = 0 (a new user) and accumulates smoothly and monotonically, so no arbitrary initial state is simulated and there is no initial transient to discard. Figure S2 summarizes the entity–attribute–activity structure of the model (ODD protocol documented in the code repository).

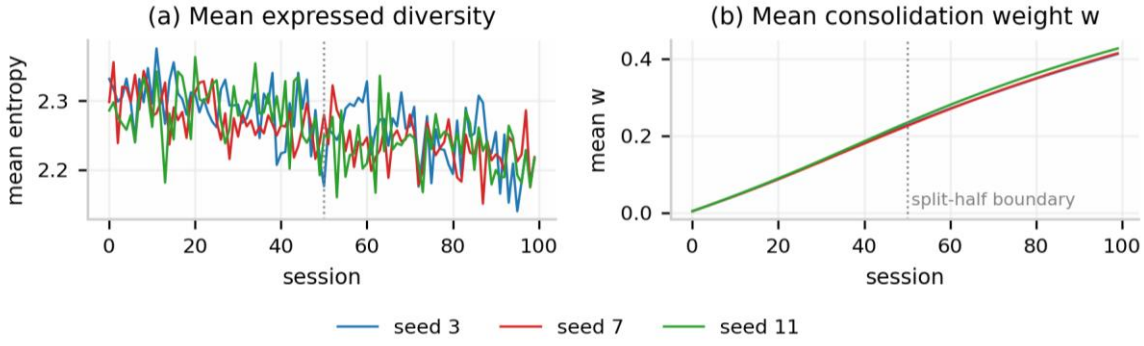


Figure S1. Slow-memory trajectories without a warm-up period (baseline world, N = 5,000; seeds 3, 7, 11). The dotted line marks the split-half measurement boundary (t = 50).

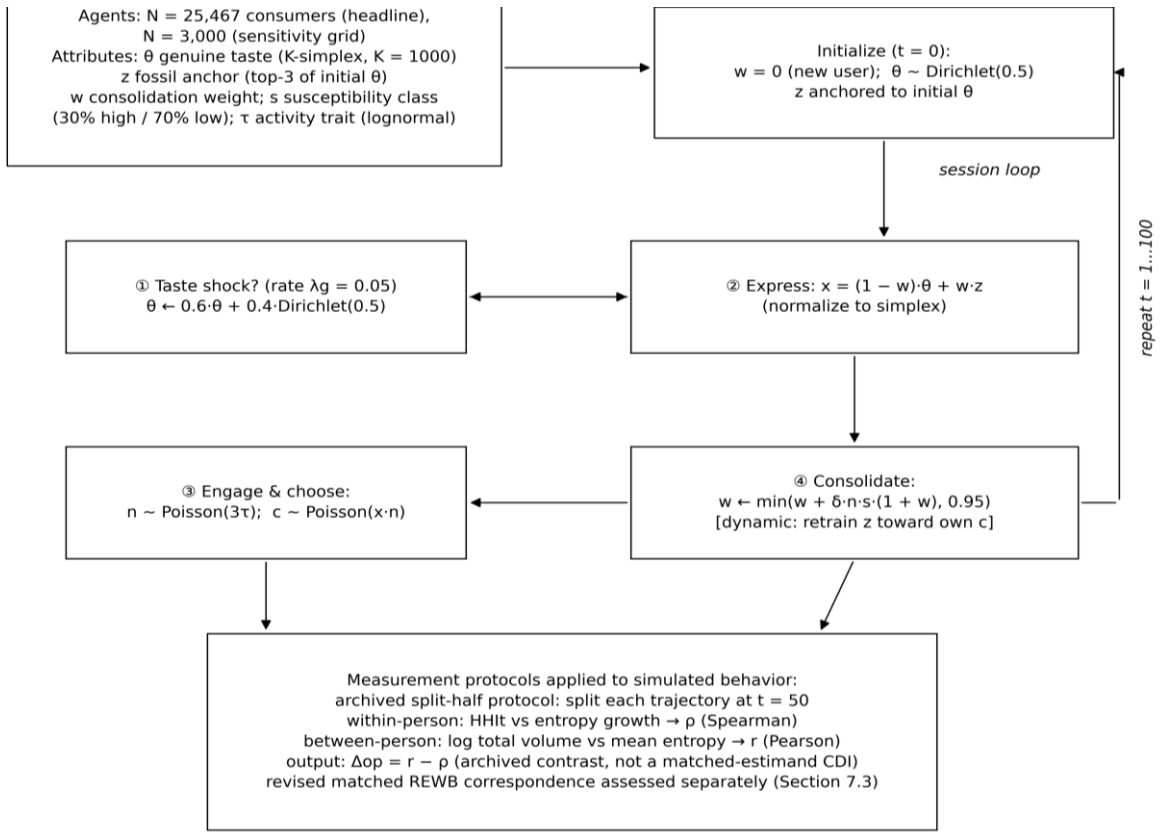


Figure S2. Structure of the slow-memory simulation: agent attributes (θ, z, w, s, τ), the per-session activity cycle (taste shock → expression → engagement/choice → consolidation), and the archived split-half measurement. The archived operationalization contrast Δop = r − ρ is not a CDI under the matched-estimand definition (Section 2.2).

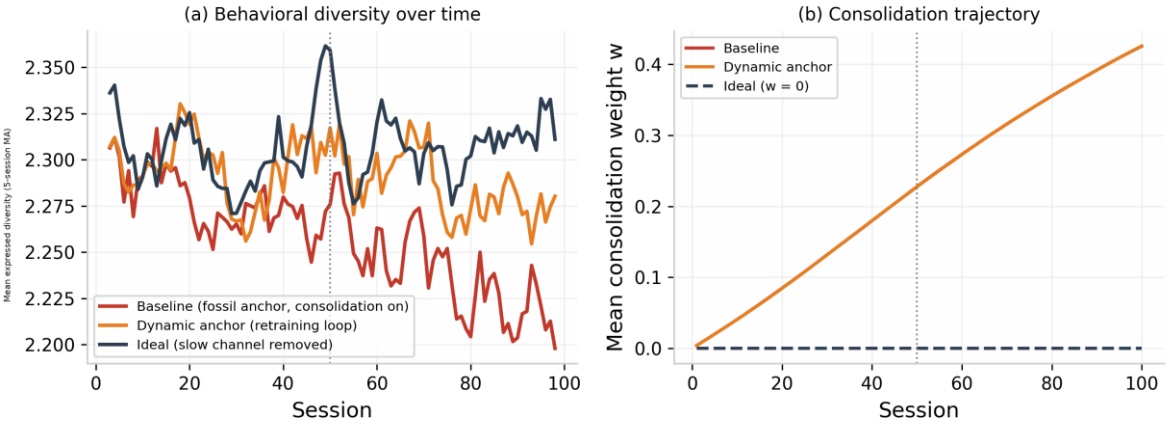


Figure S3. Simulated diversity and consolidation trajectories (N = 3,000, seed 7). (a) Mean expressed diversity (5-session moving average) under the baseline, dynamic-anchor, and ideal (no-slow-channel) conditions; (b) consolidation weight w over sessions; the dotted line marks the split-half boundary. These synthetic agents are unrelated to the archived empirical sample (N = 25,467) used only to define the legacy reference values in Table S11 and Figure S4.

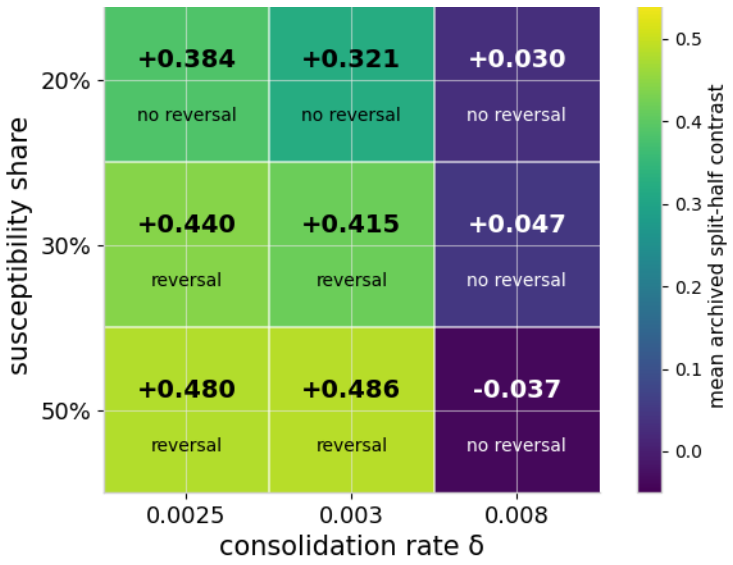


Figure S4. Archived split-half contrasts across the parameter grid.

S2.6. CDI Interpretation Boundaries

CDI is a contrast between matched caliber-specific estimates, not a standalone panel estimator. Its interpretation depends on the scale and construction of the constituent coefficients; raw CDI magnitudes are therefore not evaluated against generic effect-size thresholds. When common-scale comparison is required, the paper reports the standardized common-scale contrastCDI^*. Opposite-sign constituent estimates mechanically yield a comparatively large absolute contrast, but this algebraic property is not engaged by the focal Study 1 result because both primary coefficients are negative.

Related constructions serve orientation only and supply no numerical benchmarks for a raw coefficient contrast: Cohen's q for comparing two correlations [78], a dynamic index of technological change [39], and the concentration operationalizations used elsewhere in the paper, namely the Herfindahl–Hirschman Index [64] and the Shannon index [67]. CDI is not presented as a newly invented general-purpose estimator.

Table S10. Mapping from analytical objects to implementation artifacts.

Analytical object	Equation	Data object	Code/script	Output/table	Manuscript location
brand share / maxshare / HHI / entropy	Equations (2)–(4) (§3.3)	consumer-wave analytical panel	pipeline documented in Table S3	track1_panel-derived outputs	§3.3, §4.2
between/within components	Equations (5), (6)	per-consumer sufficient statistics	bootstrap_cdi.py	task1_report.json	§3.4
REWB/Mundlak focal model	Equation (7)	consumer-wave panel (60,423; 242,771)	bootstrap_cdi.py (dual-path verified)	results_study1_rewb.csv	§3.5, §4.3

Analytical object	Equation	Data object	Code/script	Output/table	Manuscript location
CDI /CDI^*	Equations (8), (10)	joint coefficient estimates	bootstrap_cdi.py	task1_report.json; c3_portability_master_table.csv	§3.6, §4.3
joint bootstrap / BCa	Equation (9)	consumer-cluster replicates (B = 2,000)	bootstrap_cdi.py	cdi_bootstrap_2000.csv	§3.6, §4.3
ACT1 opportunity conditioning	Equations (11), (12)	valid behavioral-event counts (code object n_click)	activity module (frozen import)	portability_multiverse_master.csv (ACT1)	§3.8, §4.4
D2 / T2a / T2b temporal constructions	Equation (13) + FE definitions	window timestamps	temporal module (frozen import)	portability_multiverse_master.csv (D2, T2a, T2b)	§3.8, §4.4
Study 2 quadratic trajectory	Equation (14)	review-level / user × tenure-year objects	Study 2 pipeline (archived)	results_study2_study3.csv	§5.2
Study 3 breadth composite and normalizations	Equations (15)–(17)	consumer-level breadth object	Study 3 pipeline (archived)	r1_item5_activity_normalized_results.csv	§6.2, §6.4
generative probe regimes	Equations (18)–(22)	synthetic consumer × window panel	sim_rewb_boundary.py	simA_results.json; simA_summary.csv	§7.2–7.4

S2.7. Archived Study 1 Split-Half Fidelity Audit (Provenance Only)

This subsection documents the computational fidelity of the retired split-half provenance analysis; it is not a validation procedure for the revised REWB CDI. The full-sample run first replicated the pilot pipeline on the identical pilot slice (the first 30,000 users in file order) to verify the frozen values before expanding to all 173,831 users. On the pilot slice (30,000 → 6860 → 4320 analytic users; tertile group sizes 1632 / 1628; tail counts 304 / 68 / 17), the frozen-value comparison passed 46/46 items (tolerance 5×10^{-4}; t-value tolerance 0.02). The SciPy and hand-written standard-library midrank implementations agreed on all 24 checks, with maximum absolute differences of 8.3×10^{-17} on the pilot slice and 1.9×10^{-16} on the full sample.Under the disclosed branch rule (same sign and $|\varrho| \in [0.08, 0.20]$), the pilot-to-full decision rule was triggered: the full-sample values reported in the main text became the estimates of record, and the pilot values are reported as pre-validation.

S2.8. Legacy Scale Calibration of Equation (1) (Provenance Only)

This section calibrates Equation (1) at the scale level; the slow-memory agent simulation of Section 7 is a mechanism-level generative sufficiency test. The calibration below is a legacy analysis benchmarked to archived-sample targets ($\rho = -0.130$, $r = +0.287$, tertile contrasts +0.415/+0.242) that have been retired from the revised evidence base; it is retained for reproducibility only and is not used as evidence for the revised CDI. The two share parameter ranges but are different experiments (different seeds and estimators), so their numerical outputs are not expected to coincide. The calibration reported in Section 2.3 simulates Equation (1) for 25,467 synthetic consumers over $T = 100$ sessions with split-half windowing mirroring Study 1. Per consumer, $x_{t+1} = x_t + \lambda_g - \delta \cdot (x_t - x_0)$, with $\log \lambda \sim N(0, 0.7^2)$ (activity heterogeneity, generating the between-person association), $\delta \sim |N(0.004, 0.0065^2)|$ (consolidation-pull heterogeneity, generating the within-

person association), starting breadth xstart ~ N(2.6, 2.0^2) truncated at 1.2, and fixed point $x_0 = 2.0$. Measurement: $H(t) = \log\ x(t)$; $\Delta H = H_2 - H_1 + \mathrm{N}(0, 0.52^2)$; the baseline consolidation proxy loads on δ with noise N(0, 0.03^2); cumulative volume $V = 100\lambda$ with log-noise (sd 0.9); mean entropy carries noise N(0, 0.8^2). At the explored reference point ($\lambda_g = 0.055$, $\delta = 0.004$) the simulation reproduces $\rho = -0.126$, $r = +0.280$, and tertile contrasts +0.365/+0.205 against the targets $\rho = -0.130$, $r = +0.287$, +0.415/+0.242. The explored simulation range — $\lambda_g \in [0.04, 0.07]$ breadth units per session and $\delta \in [0.0025, 0.008]$ per session (half-life 87–277 sessions) — is stable across seeds 42, 7, and 2026. The computation is fully deterministic; code is publicly available at the same repository (https://github.com/veranolu/C3-ABM). The empirical targets in the rightmost column are legacy estimates from the archived Study 1 analytic sample whose lineage is disclosed in Section 4.1; the calibration exercise is unaffected, because it demonstrates scale-level correspondence rather than estimation on the primary sample.

Table S11. Distinct simulation worlds and archived reference targets.

Quantity	Calibration world (§2.3, S2.8)	Agent world (Section 7)	Empirical target (Study 1, archived sample; see Section 4.1)
Within-person ϱ (Spearman)	−0.126	−0.148	−0.130
Between-person r (Pearson)	+0.280	+0.266	+0.287
Tertile contrast	+0.365 / +0.205	(held-out ordering check, Section 7)	+0.415 / +0.242

Note. The two simulation columns are independent experiments (different seeds and estimators) and are not expected to coincide; each is judged against the empirical targets in its own right. The empirical targets are legacy archived-sample estimates (lineage disclosed in Section 4.1), retained for reproducibility only and not used as evidence for the revised CDI.

S3. Full Portability, Measurement, and Auxiliary Audits

This section reports the revised Study 1 estimates introduced in Section 4.3: the same-sample correlation decomposition (Table S12), the regression-to-the-mean and baseline-coupling decomposition (Table S13), the primary-sample census chain (Table S14), the window-sensitivity grid for the panel decomposition (Table S15), and the disclosed 2 × 2 audit of the legacy anchor (Table S16). All correlations are Spearman rank correlations; bootstrap intervals use 1,000 resamples. The Study 3 activity-normalization audit of behavioral breadth is reported in Section S3.4 (Table S20).

S3.1. Archived Provenance and Primary-Panel Reconstruction Audits

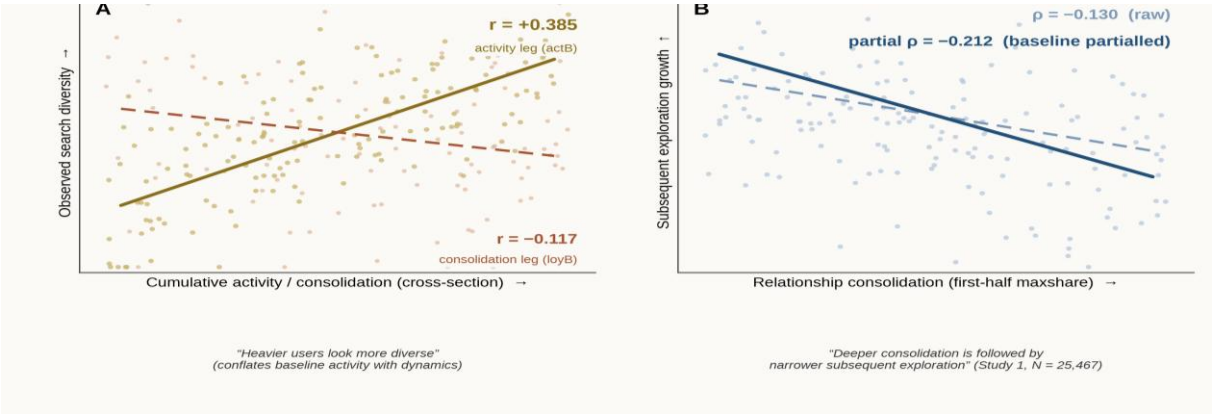


Figure S5. Archived split-half operationalization contrast (legacy analytic sample, $N = 25{,}467$). Retained for provenance; it does not support the revised consolidation claim in Section 4.

Table S12. Same-sample caliber decomposition for Study 1 ($N = 25{,}467$; Spearman rank correlations; archived analytic sample).

Predictor (first half)	Cross-sectional outcome: entropy (first half)	Longitudinal outcome: Δentropy (second − first half)
Activity (event volume)	+0.385 ($p < .001$)	−0.209 ($p < .001$)
Consolidation (maxshare)	−0.117 ($p < .001$)	−0.130 ($p < .001$)
Operationalization gap	+0.502, 95%CI [+0.485, +0.520]	—

Table S13. Regression-to-the-mean and baseline-coupling decomposition ($N = 25{,}467$; archived analytic sample).

Quantity	Estimate	Interpretation
Baseline coupling: entropy(first) × Δentropy	−0.472	Higher initial diversity mechanically declines
Activity within leg, raw / partial	−0.209 / −0.034	Largely mechanical (RTM artifact)
Consolidation within leg, raw / partial	−0.130 / −0.212	Strengthened within the archived construction; archived diagnostic only, not evidence for the revised claim
Activity × consolidation	−0.114	Weak coupling; distinct engagement facets

Note. The decomposition applies only to the retired split-half construction. Revised matched-panel measurement diagnostics are discussed immediately below and summarized in main-text Appendix Table C2.

Measurement and regression-to-the-mean diagnostics for the revised construction.

Table S13 documents regression-to-the-mean properties of the retired split-half construction. Because that design differs from the matched REWB analysis, the revised construction requires separate measurement diagnostics. Two calibrated generating models preserve the observed event-count structure and the brand-to-category measurement pipeline. Model A allows wave-specific preference parameters, whereas Model B assumes stable consumer-level preferences. Their mean pseudo-within coefficients are −2.2845 and −2.3983, respectively, compared with the empirical −3.5132; the corresponding pseudo-CDIs are −0.0084 and +0.0604, and the pseudo-within magnitudes represent 65.0% and 68.3% of the observed within-consumer coefficient, respectively. None of 100 full-panel replications per model reaches the observed CDI of +0.5174.

A nonoverlapping cross-wave design provides a complementary check. Its association is −1.2653; applying a correction based on split-half reliability of 0.427 yields approximately −2.96, compared with the same-wave estimate of −2.8588. This comparison depends on the assumptions of the reliability correction and is not a direct estimate of the latent behavioral component.

Together, these diagnostics show that the two specified measurement-generating processes do not reproduce the observed matched contrast, while the cross-wave result remains negative under a different measurement arrangement. They neither rule out all measurement-driven explanations nor establish that the full within-person magnitude is behavioral. The legacy RTM decomposition does not carry over automatically to the matched REWB estimand.

Table S14. Primary-sample census chain for the REWB panel (reconstructible end-to-end from raw data).

Step	Criterion	Users	Excluded at step
U	Parsed from raw clickstream	173,831	—
C1	≥ 20 raw behaviors	142,435	31,396
C2	Longitudinal windowing	142,434	1
C3	Identifier mapping	140,592	1,842
C4	Wave construction (≥3 valid windows; ≥3 mapped events per valid window)	60,423 (242,771 obs; mean 4.02 waves)	80,169

Note. C1–C3 exclusions (31,396 + 1 + 1,842 = 33,239 short-history users) occur before C4 and are disjoint from the C4 set; the census closes exactly: 33,239 + 0 + 80,169 + 60,423 = 173,831 (time = 0 exclusions: 0).

Table S15. Window-sensitivity grid for the REWB decomposition ($N = 60{,}423$).

Wave definition	β_W	CDI ($\beta_B - \beta_W$)	Significance
5 waves (coarse)	−3.6819	+0.7302	$p < .001$
10 waves (baseline)	−3.5132	+0.5174	$p < .001$
20 waves (fine)	−3.3505	+0.4327	$p < .001$

Table S16. Disclosed 2 × 2 audit (algorithm × sample universe) of the legacy anchor correlation.

Cell	Algorithm	Sample universe	ϱ
A	v3	v3 ($N = 142{,}434$)	+0.3634
B	v3	REWB ($N = 60{,}423$)	+0.3209
C	track-1	v3	+0.1406
D	track-1	REWB	+0.0951

S3.2. Representation-Audit Matrices (Raw and Common-Scale Standardized)

The representation audit summarized in Section 4.4 is reported in full below. All estimates use the same reconstructible panel (60,423 consumers; 242,771 consumer-window observations), the same Mundlak REWB three-column design [1] $m_i dev_{it}$, and consumer-cluster-robust standard errors; only the focal predictor changes. CDI = $\beta_B - \beta_W$, with a Wald χ^2 test of H0: $\beta_B = \beta_W$. Standardized coefficients use common full-panel SDs ($\beta_* = \beta \cdot s_P / s_Y$; $s_Y = 0.5931448$; $s_P = 0.1297659$ for maxshare, 165.6982 for n_{click}, 1.5951480 for $\log_{nclick}$, so the four cells of each column are scale-comparable. Uncertainty uses a consumer-cluster bootstrap (B = 2,000, seed 20260915) with BCa-corrected intervals.

Table S17. Representation audit, raw coefficients: between- and within-person components and CDI by predictor.

Predictor	β_B (SE)	β_W (SE)	CDI (SE)	Wald χ^2	BCa 95% CI
maxshare (consolidation)	−2.9957 (0.0194)	−3.5132 (0.0110)	+0.5174 (0.0210)	609.61	[+0.4757, +0.5573]
n_click (activity, raw)	+0.001084 (0.000037)	+0.001418 (0.000019)	−0.000334 (0.000029)	136.22	[−0.000387, −0.000278]
log_nclick (activity, log)	+0.3893 (0.0034)	+0.2432 (0.0005)	+0.1461 (0.0034)	1895.29	[+0.1396, +0.1524]

Table S18. Representation audit, common-scale standardized coefficients.

Predictor	β_B^*	β_W^*	CDIstd	BCa 95% CI
maxshare (consolidation)	−0.6554	−0.7686	+0.1132	[+0.1041, +0.1219]
n_click (activity, raw)	+0.3027	+0.3962	−0.0934	[−0.1081, −0.0776]
log_nclick (activity, log)	+1.0471	+0.6541	+0.3930	[+0.3755, +0.4099]

Bootstrap CDI distributions (B = 2,000): maxshare mean = median = 0.51808, SD 0.02063, skew +0.024, percentile CI $[0.47707, 0.55884]$; n_click mean −0.000335, SD 0.0000283; log_nclick mean 0.14605, SD 0.00332; bootstrap SDs match the analytic standard errors and BCa ≈ percentile intervals throughout. Paired contrasts on the same 2,000 resamples: Δmaxshare−raw = $+0.2066$, BCa 95% CI $[+0.1875, +0.2230]$; Δmaxshare−log = -0.2798, [−0.2998, −0.2588]; Δraw−log = -0.4864, [−0.5007, −0.4729]; all exclusion probabilities 0 or 1. Under raw activity, caliber divergence reverses direction (the interval lies entirely below zero; this is a sign reversal, not an attenuation toward zero). The log transformation produces substantial functional-form sensitivity in the activity-based CDI; no mechanism claim is attached to this cell. The standardized log-activity contrast (+0.3930) is numerically larger than the consolidation contrast (+0.1132) and is disclosed as such; interpretation is restricted to functional-form sensitivity pending variance-decomposition diagnostics.

S3.3. Specification-Sensitivity Envelope Grid

The revision-stage specification grid summarized in Section 4.4 is reported in full. The grid comprises 14 revision-stage rows (13 canonical specifications in families A–D, $K = 13$, varying the temporal construction, sample support, outcome scaling, and model specification around the primary design, plus the separately reported S-E2 construct variant) and four theory-directed extension rows (ACT1, T2a, T2b, T3), added at the evidence-integration stage, for 18 reported specifications in total. T3 is not identifiable in the released dataset. The Miller–Madow row S-C3 is a scenario-dependent sensitivity estimate, not a specification. Each family perturbs a distinct analytical choice whose influence on the portability diagnosis is itself substantively informative: family A (window length) probes observation-window sensitivity; family B (sample support) probes support-selection sensitivity; family C (outcome scaling) probes representation and functional-form sensitivity; family D (wave-position fixed effects, baseline coupling, and lagged outcome) probes temporal-construction and specification sensitivity. The construct variant S-E2 (brand HHI) probes construct-representation sensitivity. The specification grid is interpreted as a sensitivity envelope rather than as a set of independent robustness replications. The construct row E1 duplicates the primary specification and is therefore not listed as an independent row; the construct variant S-E2 alters the focal representation and is reported separately below the canonical grid. All rows use consumer-cluster-robust standard errors and a consumer-cluster bootstrap (B = 2,000, BCa intervals); S-A1 and S-A3 are imported from the frozen horizon analysis of Section 4.3. Row by row, Table S19 records 18 reported specifications—14 revision-stage rows (13 canonical specifications in families A–D plus the separately reported S-E2 construct variant) and four theory-directed

extension rows—followed by one final sensitivity row (S-C3) that is reported outside the specification count. Of the 18 specifications, 17 are estimable (T3 is not). Figure S6 displays each estimable contrast once (17 points). All 17 estimable rows exhibit $\beta_B < 0$ and β_W. Figure S7 summarizes the end-to-end analytical and reproducibility workflow linking these analyses to their scripts, outputs, and provenance documentation. Complementary evidence from Study 2, Study 3, and the generative simulation maps temporal, measurement/interpretive, and mechanistic boundaries rather than providing replications of the primary analysis. A Miller–Madow bias-corrected entropy row (S-C3) is reported as two scenario-dependent first-order Miller–Madow sensitivity estimates (a numerical interval across the two examined scenarios), not as an estimated specification or a strict mathematical bound.

Table S19. Full specification and construction audit of the CDI (18 reported specifications; T3 not identifiable in the released dataset; S-C3 is a Miller–Madow scenario sensitivity, not a specification).

Spec	Description	N (cons./obs.)	β_B	β_W	CDI [BCa 95% CI]	CDI^*
S-A1	Temporal: W = 5	39,650 / 136,152	−2.9517	−3.6819	+0.7302 [+0.6811, +0.7786]	+0.1583
S-A2	Temporal: W = 10 (primary)	60,423 / 242,771	−2.9957	−3.5132	+0.5174 [+0.4757, +0.5573]	+0.1132
S-A3	Temporal: W = 20	72,166 / 348,585	−2.9178	−3.3505	+0.4327 [+0.3951, +0.4700]	+0.0967
S-B1	Sample support: MIN_RECORDS = 20, ≥3 valid waves (primary)	60,423 / 242,771	−2.9957	−3.5132	+0.5174 [+0.4757, +0.5573]	+0.1132
S-B2	Sample support: MIN_RECORDS = 30	59,704 / 240,512	−3.0127	−3.5295	+0.5169 [+0.4745, +0.5577]	+0.1123
S-B3	Sample support: MIN_RECORDS = 40	58,351 / 236,161	−3.0325	−3.5517	+0.5192 [+0.4783, +0.5616]	+0.1120
S-B4	Sample support: ≥4 valid waves	34,174 / 164,024	−3.1287	−3.6321	+0.5034 [+0.4444, +0.5627]	+0.1073
S-C1	Outcome: raw entropy (primary)	60,423 / 242,771	−2.9957	−3.5132	+0.5174 [+0.4757, +0.5573]	+0.1132
S-C2	Outcome: normalized entropy, H/log(K) (constant rescale)	60,423 / 242,771	−0.6917	−0.8112	+0.1195 [+0.1098, +0.1287]	+0.1132
S-D1	Model: primary REWB/Mundlak	60,423 / 242,771	−2.9957	−3.5132	+0.5174 [+0.4757, +0.5573]	+0.1132
S-D2	Model: wave-position fixed effects	60,423 / 242,771	−3.0408	−2.2331	−0.8077 [−0.8509, −0.7648]	−0.1767
S-D3	Model: baseline-coupling adjustment (not causal)	60,423 / 242,771	−1.5381	−3.5132	+1.9750 [+1.9434, +2.0083]	+0.4321
S-D4	Model: lagged-outcome dynamic	60,423 / 182,348	−1.8039	−2.3300	+0.5260 [+0.4936, +0.5556]	+0.1321
ACT1	Opportunity: activity-conditioned REWB (Mundlak-decomposed log(1 + valid event count))	60,423 / 242,771	−2.6377	−2.1566	−0.4812 [−0.5278, −0.4389]	−0.1053

Spec	Description	N (cons./obs.)	β_B	β_W	CDI [BCa 95% CI]	CDI^*
T2a	Temporal: continuous elapsed months since first observation	60,423 / 242,771	−2.8510	−2.8503	−0.0007 [−0.0402, +0.0367]	−0.0002
T2b	Temporal: integer-month fixed effects (sparse support; 99.6% of observations in month 0)	60,423 / 242,771	−2.9908	−3.5094	+0.5186 [+0.4770, +0.5587]	+0.1135
T3	Temporal: calendar-period fixed effects	not identifiable (dataset property)	—	—	—	—
	Construct variant (alters the focal representation; reported separately, not counted in K)					
S-E2	Construct: brand HHI = $\Sigma\, s_j^2$	60,423 / 242,771	−3.6566	−3.7130	+0.0565 [−0.0039, +0.1171]	+0.0123
S-C3	Outcome: Miller–Madow first-order bias-corrected entropy (two assumed category-count scenarios; sensitivity range, not a preferred model or strict bound)	60,423 / 242,771	—	—	−0.0847 to −0.0077 (sensitivity range)	—

Note. Intervals are consumer-cluster bootstrap BCa 95% (B = 2,000, seed 20260915). The grid is a sensitivity envelope, not independent replications: several entries are nested, algebraically related, or share the same underlying estimate. E1 duplicates the primary specification and is not listed as an independent row. Exclusions fixed in the disclosed grid: B5 (retained share 0.279, below the disclosed 0.40 support threshold), D5 (duplicate of the parsimonious primary model), E3 (no third concentration metric exists in the frozen dataset and code lineage), and a C2 variant normalizing by observed within-window support (inadmissible by construction, as it treats the observed purchase support as the opportunity set).

Family B (sample support) is tightly clustered around the primary estimate (+0.5034 to +0.5192), and the C2 constant platform-catalog rescale leaves CDI^* invariant, as it must. The sharpest boundaries come from the temporal and opportunity rows: D2 (wave-position fixed effects) reverses the sign of the CDI (−0.8077 [−0.8509, −0.7648]), ACT1 (opportunity conditioning) likewise reverses it (−0.4812 [−0.5278, −0.4389]), T2a (continuous elapsed months) collapses the contrast to zero (−0.0007 [−0.0402, +0.0367]). T2b reproduces the focal estimate (+0.5186 [+0.4770, +0.5587]) on sparse support (241,686 of 242,771 observations in month 0; 1,070 and 15 in months 1 and 2). D3 (baseline-coupling adjustment) inflates the contrast (+1.9750 [+1.9434, +2.0083]) and is a magnitude-sensitivity boundary rather than a confirmation or a causal adjustment. D4 (lagged outcome) preserves direction and magnitude (+0.5260 [+0.4936, +0.5556]). S-E2 (brand HHI) attenuates the contrast toward zero (+0.5565 [−0.0039, +0.1171]; $CDI^* = +0.0123$), an exception showing that concentration measured across the full distribution does not support a cross-caliber divergence. T3 is not identifiable in this dataset: 98.84% of consumers enter at the first calendar period, and the recoded per-consumer relative-time field has a maximum origin offset of 934 seconds (median span between first and last retained waves 3.9 days, versus a median full activity span of 30.7 days). A permutation check (B = 1,000; P = 0.001) validates the estimation and bootstrap pipeline against a shuffled-label null and is reported as pipeline validation, not as a causal placebo.

The permutation diagnostic referenced in Section 4.4 randomly reassigns each consumer's wave-level outcome–predictor pairing within that consumer's own sequence (1,000 permutations), preserving the between-consumer structure of the measures while destroying genuine within-wave covariation. None of the 1,000 permuted cross-caliber

contrasts reached the observed value, yielding P = 0.001. The test is a pipeline-level diagnostic of mechanical coupling, not an independent identification result. Because the permutation holds each consumer's mean fixed, the between component is bit-invariant across permutations ($SD = 0$) and the permuted contrast distribution centers on the between association rather than on zero; the calibrated measurement-generating and cross-wave reliability diagnostics reported with Table S13 supply the separation this test cannot provide on its own.

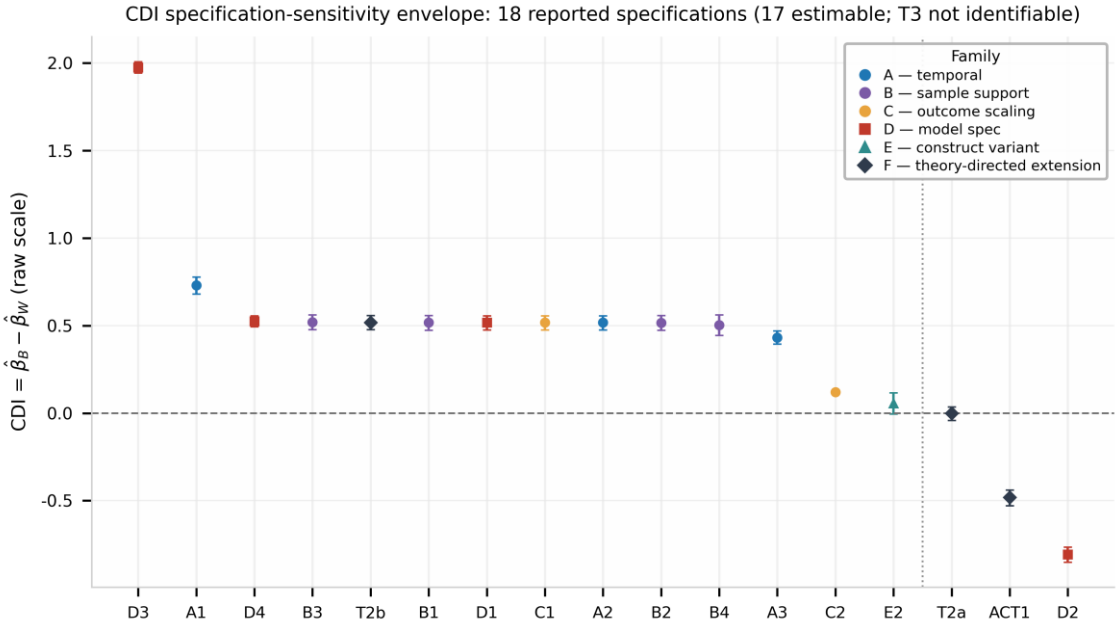


Figure S6. CDI across analytical constructions. Points show matched-estimand contrasts with consumer-cluster bootstrap BCa 95% intervals (B = 2,000) for the 17 estimable specifications; T3 is not identifiable and carries no point. Estimates appear in Table S19.

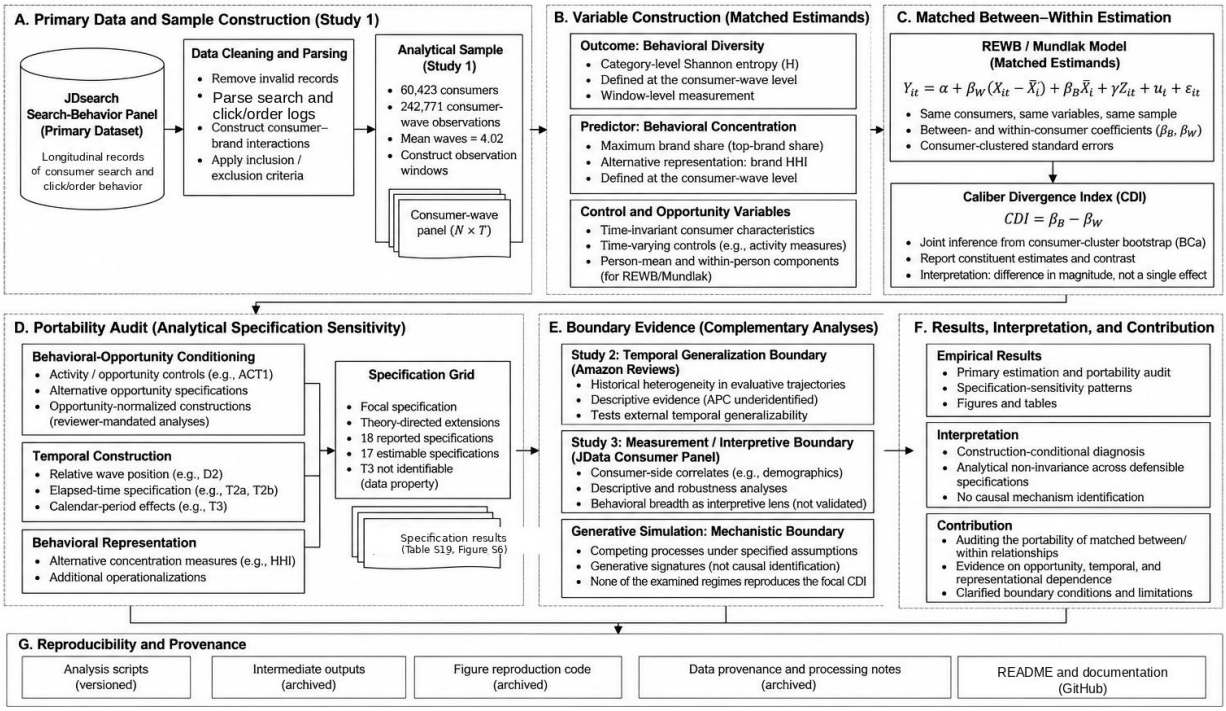


Figure S7. Analytical and reproducibility workflow. The diagram traces the evidence chain from data and sample construction through estimation and the portability audit to analysis scripts, outputs, and provenance documentation; it does not imply causal identification.

S3.4. Activity-Normalized Behavioral-Breadth Audit (Study 3)

The activity-normalization audit summarized in Section 6.4 is reported in full below. The raw behavioral-breadth composite is diversity_score = log(1 + n_unique_cates) + log(1 + n_unique_brands), computed on the Study 3 age-valid sample ($age > 0$, $N = 23{,}644$); the gender contrast uses the subset with valid male/female codes ($N = 13{,}081$). Two preselected normalizations of behavioral opportunity were evaluated: (A) a ratio construction, norm_div_ratio = diversity_score / log(1 + n_actions). (B) A residual construction, described operationally per the archived analysis record: the OLS residual of diversity_score on log(1 + n_actions), fitted on the full age-valid sample, with the gender contrast inheriting that fit ($d = +0.2519$). Refitting the residual construction within the gender subset instead yields $d = +0.2514$. Associations are tested with Spearman ϱ (age, consumption level) and Welch t with Cohen's d (gender). Full precision is archived in r1_item5_activity_normalized_results.csv; the table reports display values.

Consumption-level group means of the ratio-normalized composite rise from 0.630 to 0.686 across levels 1–5 (0.630 / 0.615 / 0.631 / 0.653 / 0.686; $N = 25$ / 1,281 / 4,665 / 7,812 / 9,861), preserving the shape of the raw-composite gradient (3.857 / 3.587 / 3.769 / 4.067 / 4.423). Categories per action is an additional, non-prespecified construction that changes the interpretation of breadth: the age and consumption-level associations change direction under it, whereas the gender difference attenuates to near zero and is not statistically distinguishable from zero ($d = -0.012$, $p = .539$). Behavioral-breadth associations are therefore sensitive to the specific construction of opportunity normalization; the two preselected normalizations are mutually consistent, but no unique corrected indicator exists.

Table S20. Behavioral-breadth associations under alternative opportunity constructions (Study 3).

Construction	Association	Statistic	Value	p / 95% CI	N	Note
Raw (frozen anchor)	Age	Spearman ϱ	−0.0407	$p = 3.8 \times 10^{-10}$	23,644	Frozen anchor (Sec. 6.3)
A) Ratio	Age	Spearman ϱ	−0.0218	$p = 8.2 \times 10^{-4}$	23,644	Pre-selected
B) Residual	Age	Spearman ϱ	−0.0226	$p = 5.2 \times 10^{-4}$	23,644	Pre-selected
Raw (frozen anchor)	Consumption level	Spearman ϱ	+0.2861	$p < .001$	23,644	Frozen anchor (Sec. 6.3)
A) Ratio	Consumption level	Spearman ϱ	+0.2013	$p = 1.2 \times 10^{-214}$	23,644	Pre-selected
B) Residual	Consumption level	Spearman ϱ	+0.2028	$p = 5.9 \times 10^{-218}$	23,644	Pre-selected
Raw (frozen anchor)	Gender (M − F)	Welch t (df)	11.93 (2960.8)	$p = 4.5 \times 10^{-32}$	13,081	Frozen anchor (Sec. 6.3)
A) Ratio	Gender (M − F)	Welch t (df)	10.26 (2944.0)	—	13,081	Pre-selected
B) Residual	Gender (M − F)	Welch t (df)	10.42 (2968.8)	—	13,081	Pre-selected
Raw (frozen anchor)	Gender (M − F)	Mean difference	+0.2776	[0.2319, 0.3232]	13,081	Frozen anchor (Sec. 6.3)
A) Ratio	Gender (M − F)	Mean difference	+0.0288	[0.0233, 0.0343]	13,081	Pre-selected
B) Residual	Gender (M − F)	Mean difference	+0.1664	[0.1351, 0.1977]	13,081	Pre-selected
Raw (frozen anchor)	Gender (M − F)	Cohen's d	+0.289	—	13,081	Frozen anchor 0.2892 (Sec. 6.3)
A) Ratio	Gender (M − F)	Cohen's d	+0.250	—	13,081	Pre-selected
B) Residual	Gender (M − F)	Cohen's d	+0.252	—	13,081	Pre-selected
Categories-per-action	Age	Spearman ϱ	+0.016	$p = .0137$	23,644	Alternative construction; completeness/QA disclosure, not pre-registered
Categories-per-action	Consumption level	Spearman ϱ	−0.0554	$p = 1.6 \times 10^{-17}$	23,644	Alternative construction; completeness/QA disclosure, not pre-registered

Construction	Association	Statistic	Value	p / 95% CI	N	Note
Categories-per-action	Gender (M − F)	Welch t; Cohen's d	−0.61; −0.012	$p = .539$ (ns)	13,081	Alternative construction; completeness/QA disclosure, not preregistered

Note. The categories-per-action rows are an additional, non-prespecified construction disclosed for completeness (Section S3.4).

References

Bracketed citations in these Supplementary Materials—[64], [78], [39], [67], [70], and [71]—refer to the numbered reference list of the main manuscript; no separate supplementary reference list is introduced.