THE UNIVERSITY OF CHICAGO

The Geometry of Belief Change: Diagnosing and Estimating Graded Asymmetries in Two-Wave Repeated Measures

By

Joseph Healy Epstein

July, 2025

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts Program in the Social Sciences

Faculty Advisor: James Evans

Preceptor: Jingyuan Qian

Abstract

Observed changes in survey responses appear to be shaped, in part, by the structural characteristics of the instrument designed to measure them. In two-wave panel surveys, the distribution of responses in the second wave is conditioned on participants' initial responses. Deviations from the scale midpoint at baseline impose asymmetric constraints on subsequent responses. The reduced proximity to one boundary and increased proximity to the other bias the response space, concentrating probability mass toward the farther extreme. Even after normalizing each observed change by the maximum possible displacement from the baseline response, consistent asymmetries in both the direction and magnitude of change persist across three separate two-wave panel datasets ($n = 3{,}427$; $N_{obs} = 70{,}514$). Responses tend to move away from the nearest boundary, with displacement magnitude increasing as the baseline response approaches a scale endpoint. These asymmetries persist beyond what can be accounted for by random error, regression to the mean, or sensitivity to between-wave evidence. We term this systematic tendency *open-movement bias*: A pattern in which follow-up responses disproportionately shift toward the side of the scale with greater residual support—that is, the direction offering a greater range for movement from the respondent's initial position on a bounded, one-dimensional scale. This observation motivates the core theory proposed in this thesis: that respondents' psychological tendencies interact with the structural affordances of the bounded response scale. The correlation between baseline responses and evidence-discrepancy terms remains directionally invariant, irrespective of between-wave evidence placement, due to their shared dependence on the baseline measure. This invariance reveals structurally embedded asymmetries in potential response change imposed by the bounded scale, and empirical data suggest that respondents are sensitive to these asymmetries. This thesis shifts the analytical focus from what survey responses *reveal* about latent states to how the geometry of the measurement instrument *shapes* those responses. In doing so, it advances a methodological framework at the intersection of behavioral science, measurement theory, and survey methodology.

The Problem

Most panel models—whether latent change models, fixed effects, or difference scores—treat ordinal scales as if they are linear and equidistant, implicitly assuming equal potential for upward and downward movement at every point. Common adjustments for baseline effects—such as regression-to-the-mean controls, Tobit corrections, or bounded growth models—tend to treat the bounded nature of the response scale as a nuisance to be corrected after the fact, rather than as an intrinsic feature of the data-generating process. These approaches assume symmetric variability and, by relying on post hoc adjustments or treating asymmetries as statistical artifacts, fail to account for the asymmetry introduced at the moment of response elicitation.

For example, a respondent positioned at the lower bound of a $[0,5]$ scale has only upward latitude for change, while a respondent at the midpoint retains symmetrical potential to move in either direction—but within a limited range, since the midpoint is equidistant from both bounds. This asymmetry in permissible change is a basic arithmetic property of any point that deviates from the center of a bounded, unidimensional scale. While simple, it directly challenges the common assumption of uniform change potential embedded in many conventional measurement models and their associated adjustments.

To dismiss such effects as mere "error" is to overlook a more fundamental issue: that the measurement instrument *may shape the very outcomes it is designed to capture.* Given that asymmetries in potential change are structurally embedded in the response scale following a first-wave response, it is plausible that respondents are sensitive to both the perceived and actual residual latitude available at the time of the second measurement—that is, the remaining space for movement in a given direction. Their subsequent responses may, in part, reflect this sensitivity to the remaining space for directional movement.

Inferential claims such as "respondents became more convinced" are conditioned by baseline positions, which constrain the direction and magnitude of feasible change on bounded response scales. Because movement potential varies with initial placement—independent of between-wave

evidence or intervention—such structural asymmetries complicate the detection of genuine attitudinal change and risk conflating measurement artifacts with the underlying construct. This challenge motivates a research agenda aimed at identifying geometric asymmetries and developing models that estimate instrument–response interactions, enabling clearer distinctions between ”true” opinion change and scale-induced deviation (i.e., response-measurement entanglement).

Approach, Postulates, & Proceedings

We advance a plausible and intuitive hypothesis: response shifts on bounded scales are jointly shaped by two forces—(i) structural constraints inherent to the scale, and (ii) psychological factors, including cognitive-affective states and individual dispositions. To our knowledge, the interaction between these forces has received limited empirical and theoretical attention.

We observe that both the evidence discrepancy—defined as the difference between a respondent's baseline and the between-wave anchor—and the threshold distance—the difference between that baseline and a fixed scale point (e.g., the 25th or 75th percentile)—are linear functions of the baseline value. As affine transformations of the same variable, they co-vary perfectly, producing a deterministic correlation of $\pm$, with the sign determined solely by the relative placement of the anchors. This correlation does not reflect respondent behavior or interpretation, but rather a structural artifact of variable construction—an instance of the Oldham–Lord paradox, in which spurious associations arise from shared components in difference scores.

To break this mechanical dependence, we re-express responses in terms of each respondent's maximum feasible movement. While the baseline still influences the rescaled variables, it no longer functions as a shared additive component, thereby eliminating the induced correlation. This transformation yields a diagnostic statistic, $\hat{\rho}$, which quantifies response change relative to each respondent's maximum possible displacement. This accounts for the fact that absolute movement potential is not uniform: it increases near boundaries and diminishes at the midpoint. Remarkably, even after normalizing for individual movement capacity and accounting for evidence polarity, we

continue to observe a consistent asymmetry.

Respondents near the boundaries of the scale—despite requiring larger raw shifts to achieve equivalent normalized scores—still exhibit greater normalized changes than those starting from more central positions. While normalization reduces baseline-related variation in change magnitude, it does not eliminate it. Additionally, directional asymmetries persist consistently in both raw and normalized metrics.

Analyzing three two-wave surveys ($>$ 70,000 observations), we find that these patterns in trajectory cannot be explained by classic regression-toward-the-mean (RTM): Lord–Novick tests, permutation tests, and bootstrapped inflection-point analyses reject the RTM benchmark in each panel.

Moreover, although these patterns may superficially resemble classical RTM, they differ in two key respects. First, they are quasi-deterministic—unlikely to arise from stochastic processes alone (barring the implausibility of 70,000 random survey responses). Second, they exhibit position- and group-specific reversion tendencies, rather than convergence toward a central tendency.

Complementing the normalization metric, we introduce $\tilde{\rho}$, a comparative measure that captures individual-level change relative to all other items aside from the focal one. We implement $\tilde{\rho}$ within, what we term, the *Piecewise Linear Regression Corridor* (PLRC), a new model estimating the extent to which proximity to a scale boundary at baseline influences subsequent response trajectories.

Before entering into the main body of the argument, we provide a compact preview of some recurring definitions and models.

Preliminaries

- j: individual
- i: item
- T^1: response wave 1

- T^2: response wave 2
- ΔT: response change $T^2 - T^1$
- UB, LB: Upper Bound, Lower Bound

Normalized Change index:

$$\hat{\rho}_{ji} = \frac{\Delta T_{ji}}{\max(UB - T^1_{ji},\, T^1_{ji} - LB)} \tag{1}$$

This normalization—realized individual change divided by the maximum feasible shift from baseline—recurs in various structural forms throughout the analysis. A second index introduced here is the Relative Responsiveness Index.

Relative Responsiveness index:

$$\tilde{\rho}^{\text{loo}}_{ji} = \frac{1}{I-1} \sum_{k \neq i} \frac{|\Delta T_{jk}|}{\max(|s - T^1_{jk}|,\, \varepsilon)} \tag{2}$$

where:

- $\Delta T_{jk} = T^2_{jk} - T^1_{jk}$: the change in response for individual j on item k
- $|\Delta T_{jk}|$: the absolute value of the response change (i.e., the magnitude of change)
- $|s - T^1_{jk}|$: the absolute distance between the baseline response and the evidence-aligned target s, introduced between waves 1 and 2
- $\max(\cdot, \varepsilon)$: a small-sample safeguard to prevent division by zero when the baseline equals the target
- I denotes all response items within the survey
- $\sum_{k \neq i}$: a leave-one-out (loo) summation that excludes the focal item i

This expression provides a normalized, leave-one-out average of absolute changes across all items other than i, scaled by the feasible movement in the direction of the between-wave evidence,

based on the respondent's baseline response for each respective item. It serves as a *responsiveness index* for individual j, benchmarked against their behavior on non-focal items. Next we introduce the PLRC.

PLRC:

$$\Delta T_{ji} = \begin{cases} \alpha + \beta_L (L - T_{ji}^1) + \varepsilon_j & \text{if } T_{ji}^1 < L \\ \alpha + \varepsilon_j & \text{if } L \leq T_{ji}^1 \leq U \quad \text{for } T_{ji}^1 \in [LB, UB] \\ \alpha + \beta_U (T_{ji}^1 - U) + \varepsilon_j & \text{if } T_{ji}^1 > U \end{cases} \tag{3}$$

The PLRC predicts the change in response, $\Delta T_{ji} = T_{ji}^2 - T_{ji}^1$, as a function of the baseline score T_{ji}^1 and its position relative to two internal thresholds, L and U, where $LB < L < U < UB$. These thresholds define a *neutral band*, $[L, U]$, within the bounded response scale $[LB, UB]$. Structurally induced deviations are modeled only when the baseline falls outside this corridor—that is, when $T_{ji}^1 < L$ or $T_{ji}^1 > U$. The residual term ε_j captures unexplained individual-level variance, assumed independent and identically distributed: $\varepsilon_j \sim N(\theta, \sigma^2)$. Having introduced these components, we now turn to the main exposition.

Overview

The architecture of the scale meaningfully shapes response trajectory.

Reconceptualizing the Response Scale as Environment

A substantial body of research in ecological psychology and decision science demonstrates that behavior emerges from dynamic, reciprocal interactions between organisms and their environments (Gibson, 1979; Warren, 1984; Turvey, 1992; Glenberg, 1997; Proffitt, 2006; Chemero, 2009). It is also well-established that the design and administration of surveys significantly influence the attitudes and behaviors reported by respondents (Tourangeau, 1984; Sudman et al., 1996; Groves et al., 2009; Meyer, 2015). Several well-documented phenomena arising from respondent–instrument interaction include response-order effects (Sanjeev & Balyan, 2014; Höhne & Lenzner, 2023),

question-order effects (Rasinski, et al., 2012), and anchoring biases (Tversky & Kahneman, 1974). These effects are not attributable to any individual respondent; rather, they emerge systematically from the interaction between respondents' internal processes and the format, sequence, or structure of the survey instrument.

Indeed, elements of survey construction—such as item wording, layout, and response options—can meaningfully influence outcomes. Whether these instrument-related influences should be treated as "error" is disputed. Analysts may treat such features as sources of nuisance to be statistically controlled or sources of substantive meaning to be interpreted as components of the response process (de Boeck & Wilson, 2004).

When the latter view is adopted, it can be interpreted within an ecological perspective (Barker, 1968). Extending these effects, we propose that respondents internalize not only question content and contextual cues (Kalton & Schuman, 1982), but also deeper structural properties of the instrument itself.[1] We conceptualize responses as jointly shaped by respondent dispositions and the formal properties of the measurement instrument.

Drawing on Gibson's theory of affordances (1966, 1977, 1979/1986), we argue that bounded, unidimensional response scales present structured cues—affordances—that guide expression. Under this formulation, response options are not neutral conduits but environmental constraints. This interpretation also aligns with a constructivist perspective (Bourdieu, 1990), wherein the response scale—rather than the social structure—functions as both an external constraint and an internalized interpretive framework.

In this context, observed change may reflect not only shifts in the underlying construct, but also patterned responses to the structural affordances embedded within the measurement scale. *The metric used to assess change may, in part, shape the very responses it is intended to quantify.*

While foundational work on scale effects and cognitive response processes has been well established (e.g., Tourangeau, Rips, & Rasinski, 2000), our contribution introduces a temporal

[1]This claim does not imply mutuality between agent and environment, as respondents do not alter the structure of the scale.

structural dynamic—specifically in the domain of change scores—that remains under-explored in existing survey methodology.

We contend that the bounded structure of response scales creates asymmetric opportunity spaces, which systematically constrain the direction and magnitude of within-subject change. These structural asymmetries can bias both the empirical distribution of response updates and the interpretation of their meaning. Apparent counter-evidential or "baseless" updates may thus emerge, at least in part, as artifacts of the response format rather than as expressions of genuine resistance to (or acceptance of) new information. Crucially, this argument does not concern statistical artifacts, such as those arising from correlations between initial scores and change (e.g., the Lord-Oldham paradox), but rather structural constraints inherent in the measurement instrument itself. In this respect, we extend existing survey methodology by demonstrating how response scale design can systematically influence observed patterns of individual-level change.

The Limits of Self-Report as Direct Measurement

Surveys provide a structured means of eliciting information from a sampled subset of individuals (Scheuer, 2004), forming the empirical basis for inferring population-level dynamics from individual responses (Haaland et al., 2025; Celhay, Meyer, & Mittag, 2024). However, the assumption that self-reports offer a direct window into internal states—such as attitudes or beliefs—is overly simplistic.

This assumption presumes respondents can access and articulate their cognitive and affective states accurately, unaffected by context, memory, mood, or social desirability. Even if such self-access were possible, two barriers remain: the constraints imposed by the response medium[2], and the interpretive lens through which responses are received.

As for the medium, surveys often employ fixed-response formats that shape the very expressions they seek to elicit. Response options constrain both the content and form of participants' expressions,

[2]By "medium," we refer to the response format—such as fixed-choice or open-ended—not the mode of administration.

potentially shaping how they interpret the prompt itself. Such constraints can subtly redirect the response process—from an ”open retrieval” of beliefs or attitudes to a bounded selection among predefined alternatives. As a result, participants may adapt their interpretations of the question to align with the available response options (Tourangeau et al., 2000, p. 191). This potential alteration can blur the distinction between what the respondent thinks and what the instrument permits them to say, introducing ambiguity into the mapping between internal states and recorded responses.

On the interpretive side, respondent misinterpretations—such as conflating debt with deficits or correlation with causation—are often attributed to bias or flawed reasoning (Stantcheva, 2022). Yet, in the absence of standardized interpretive benchmarks, such responses may reflect ambiguity in item design rather than ”irrationality.” A common example of this interpretive gap is the treatment of complex or contested constructs—such as freedom, welfare, or fairness—as if they were unambiguous, without verifying how respondents understand them. Such assumptions exemplify a broader class of researcher-side interpretive errors that can distort the reception and analysis of survey responses.

While rigorous instruments and protocols can mitigate measurement error (Singleton & Straits, 2009), a systematic understanding of its sources and structure remains limited. Empirical research on survey error is relatively rare, in part because researchers typically lack access to objectively verifiable “true” values against which to assess responses (Celhay, Meyer, & Mittag, 2024). This challenge is especially acute in the study of latent constructs—such as attitudes or psychological traits—which, unlike observable outcomes like income or criminal records, lack external reference points (Stantcheva, 2022).

Despite the potential for error at every stage of a single-wave survey—from question design to statistical analysis (Smaldino & McElreath, 2016)—and the absence of verifiable ground truths, researchers and social scientists continue to pursue models that capture within-person change in latent constructs over time, as part of the broader effort to advance our understanding of individual and social dynamics.

Panel Surveys

The two-wave panel survey—a minimalist empirical method for examining belief dynamics—tracks the same individuals at two distinct time points. Like most survey methodologies across disciplines, panel studies typically rely on unidimensional scales to capture responses. However, modeling response change using a unidimensional survey scale inherently imposes a path-dependent structure. Once respondents are positioned at a specific point on such a scale, any subsequent change is interpreted as linear movement along a fixed axis. Consequently, most two-wave panel designs are inherently limited in their ability to capture the trajectory of construct change, as the linear and bounded structure of the response scale constrains both the representation and interpretation of observed change. Their principal analytical value, we contend, lies in identifying deviations from stasis and assessing directional asymmetries—that is, the extent to which change is more likely to proceed in one direction than the other, independent of intervening evidence or indeed absent intervening evidence.

Although two-wave panel designs may appear straightforward they introduce additional layers of potential error. Compared to single-wave surveys, they are more susceptible to compounded measurement error, as inaccuracies from each wave can accumulate and inflate observed change scores (Amaya, Biemer, & Kinyon, 2020). Additionally, factors such as respondent recall limitations and evolving external conditions further undermine the reliability of measured change. Moreover, even when internal and external conditions are held constant, *the measurement instrument itself* can give rise to response variation absent any clear psychological or contextual basis—an effect we will demonstrate throughout this exposition.

Data Sources and Analytical Overview

The subsequent analysis draws on secondary data from the following two-wave panel studies:

Study (renamed)	*N* Participants	Original Scale	Brief Description	Event Intervention	*N* Potential Response Changes
1. Political Beliefs					
(Vlasceanu et al., 2021)	1,777	0 (inaccurate)–100 (accurate)	36 political propositions (12 neutral, 12 Democrat-leaning, 12 Republican-leaning)	18 supportive and 18 refutative facts presented between waves regarding each respective proposition	63,972
2. Health Perceptions					
(Bearth & Siegrist, 2020)	1,223	1 (no fear)–7 (high fear)	Five items on COVID-19 risk perceptions (e.g., fatality, hospital strain)	N/A. Passage of time (two months) during the SARS-CoV-2 pandemic	6,115
3. Beliefs: Religiosity & Life Outcomes					
(Anglin, 2019 – Study 2)	427	1 (negative outcomes)–9 (positive outcomes)	One two-wave response rating on religiosity and life outcomes	Randomly assigned to religion-enhancing or -disparaging between-wave evidence conditions	427

Table 1: Overview of Two-Wave Survey Datasets and Between-Wave Interventions Utilized in Secondary Analyses. We are grateful to Madalina Vlasceanu for providing access to the full dataset and thank the original authors for making their data publicly available via the Open Science Framework. All analyses presented herein are methodologically independent of the original studies from which the data were drawn. Any errors or misinterpretations are solely our own. This study neither endorses nor disputes the original findings; the datasets are employed exclusively for their empirical utility and their relevance to the objectives of the present research.

Hyperlinks to the original studies are provided in Table 1. Each dataset is summarized in its respective section, including hypotheses, evidentiary structure, and between-wave design, with additional detail in the Appendix. The structure of the paper is as follows: Section 1 provides a conceptual foundation and introduces the normalized change index, $\hat{\rho}$, as a practical diagnostic tool. Section 2 outlines the theoretical framework, emphasizing asymmetries in response space and their implications for inference. It also benchmarks the PLRC model against four alternatives and applies it to Study 1—the largest dataset, which employs a quasi-continuous Visual Analog Scale

(VAS). Sections 3 and 4 extend the analysis to Likert-based response formats using Studies 2 and 3, respectively. Section 5 discusses limitations, and Section 6 offers concluding remarks.

1. Conceptual Framing

Figures 1 through 4 illustrate a two-wave response process using randomly simulated response trajectories, providing a conceptual foundation for analyzing within- and between-subject change herein.

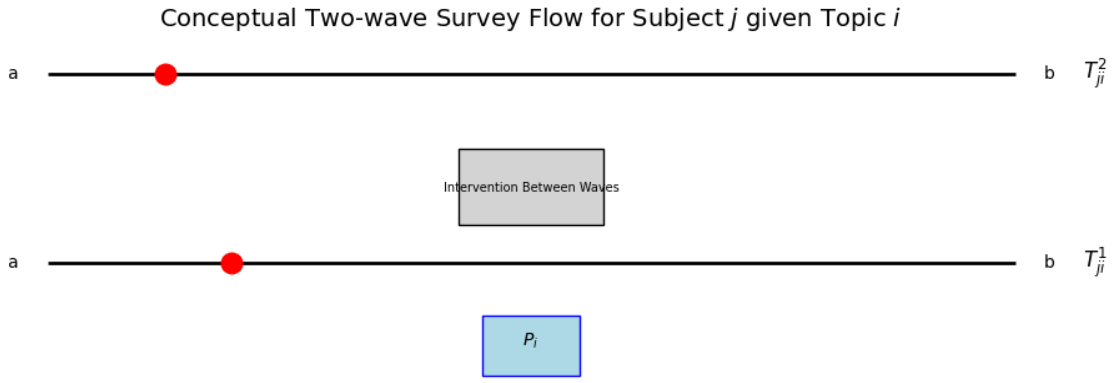


Figure 1: The subject j is presented with a proposition or question P_i. The agent records a first-wave response, T_{ji}^1, on a discrete interval scale $[a,b]$ with equally spaced integer ticks. Following an intervention—such as new evidence, a treatment, or simply the passage of time—the subject is asked to re-evaluate the same proposition P_i, producing a second response, T_{ji}^2, on the same $[a,b]$ scale.

Next, consider a linear transformation that maps the original response interval [a,b] onto a standardized scale [0,100], preserving the respondent's initial value T_{ji}^1 value. Given this transformation, what is the expected distribution of change scores—and, by extension, the expected distribution of the second-wave response T_{ji}^2?

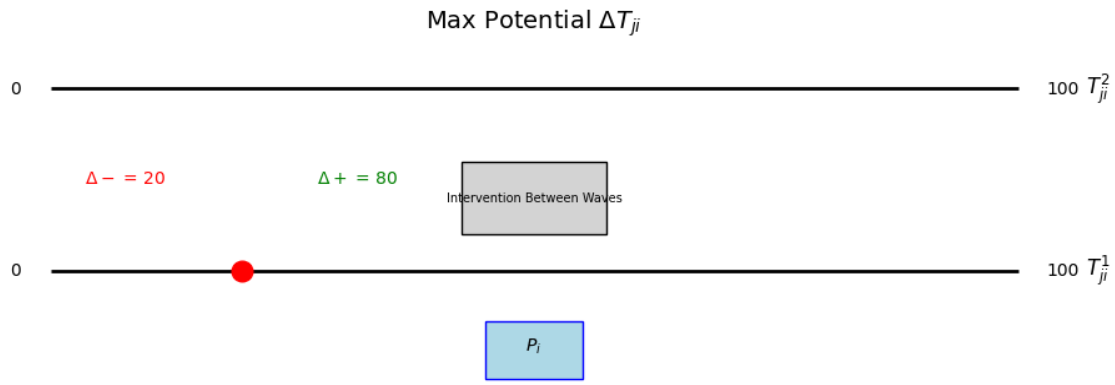


Figure 2: When the initial response is fixed at $T_{ji}^1 = 20$, the respondent has 100 alternative values on the discrete 0–100 scale, excluding the baseline. Since change is defined as $\Delta T = T^2 - T^1$, distribution of possible change scores is inherently shaped by the position of T^1 within the bounded scale. For instance, at $T^1 = 20$, more response options lie above than below, biasing change upward if T^2 is assumed to be uniformly distributed. In practice, however, response distributions seldom conform to uniformity.

Given an initial response T_{ji}^1, the second-wave response T_{ji}^2may remain unchanged or shift to one of the remaining $n-1$ values on the scale. By construction, change ΔT is constrained by the location of T^1 within the scale. On a unidimensional scale with uniformly spaced intervals, each initial position defines a distinct ”opportunity space”—that is, a finite set of allowable changes. While actual responses are not necessarily uniformly distributed, the structure of the scale implies that the number of permissible upward (or downward) shifts increases as the initial value moves closer to the lower (or upper) bound. Likewise, the maximum attainable magnitude of change increases as the initial position moves away from the scale’s midpoint.

To account for the asymmetric opportunity space, we employ Equation 1, which defines the normalized change index $\hat{\rho}$.

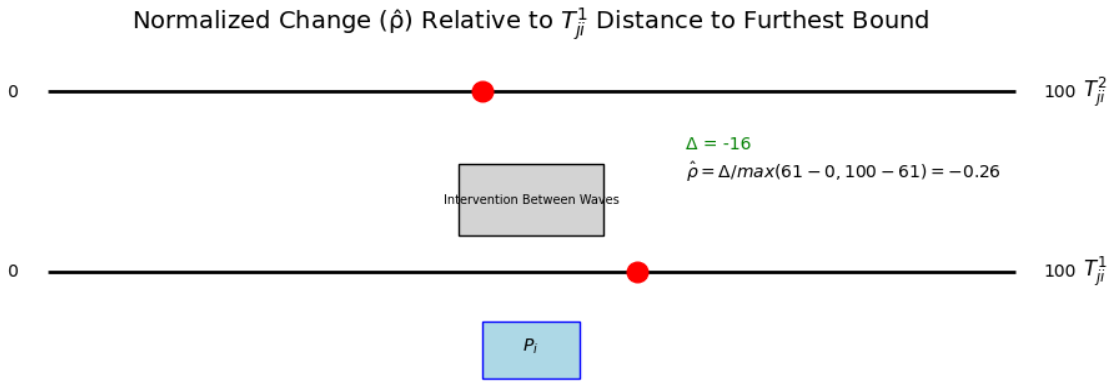


Figure 3: This formulation normalizes raw change relative to the feasible range defined by the respondent's initial position, constraining values to the interval [-1, 1]. Positive values indicate movement toward the upper bound (100), negative values toward the lower bound (0), and values near zero denote minimal change relative to the admissible range from T_{ji}^1.

By normalizing raw change relative to the feasible range of directional movement from the respondent's initial position, the normalized change index $\hat{\rho}$ corrects for structural asymmetries that are typically overlooked in standard panel survey analysis. Its bounded formulation enables interpretable estimates of both the direction and magnitude of change, grounded in the geometry of the response scale. Shortcomings of the index are discussed in the Limitations Section.

To extend the analysis, we introduce an evidence-congruency parameter that captures whether the direction of change aligns with the polarity of the intervening evidence. In Studies 1 and 3, between-wave interventions present information that either supports or challenges respondents' initial positions, thereby inducing directional pressure toward one end of the scale. To capture this, we define the evidence-congruency normalized change index $\hat{\rho}$ as follows:

$$\hat{\rho}_{ji} = \frac{|\Delta T_{ji}|}{\max(100 - T_{ji}^1,\ T_{ji}^1 - 0)} \cdot (E_i \cdot \Delta T_{ji})$$

The sign of $\hat{\rho}_{ji}$ now reflects the alignment between the direction of change ΔT_{ji} and the binary polarity of the intervening evidence E_i {-1, +1}. A positive value indicates a congruent update—change in the direction favored by the evidence; a negative value indicates an incongruent update—change counter to the evidentiary direction; and zero indicates no change.

- $\hat{\rho}_{ji} > 0$: congruent update (aligned with the direction of evidence)
- $\hat{\rho}_{ji} < 0$: incongruent update (opposes the direction of evidence)
- $\hat{\rho}_{ji} = 0$: no update

Figure 4 illustrates this using response updates from two hypothetical participants, j and l, in relation to the same proposition P_i.

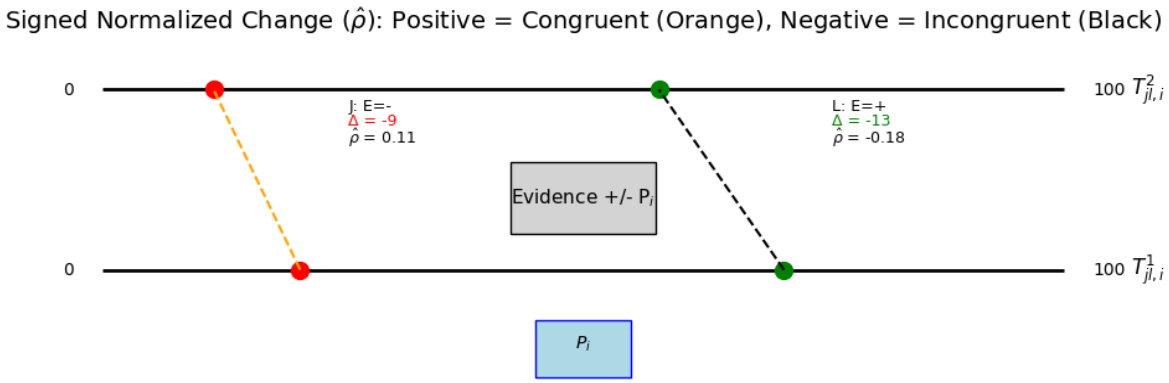


Figure 4: Incongruent changes—where positive evidence accompanies negative change (as with agent l), or negative evidence accompanies positive change—are indicated by the black dashed line. Congruent changes—where the direction of evidence aligns with the direction of change (as with agent j)—are shown by the orange dashed line.

With the conceptual foundation in place, we now turn to secondary empirical analysis. Under the definition of evidence congruency outlined above, a rational updating process would predict predominantly positive values of $\hat{\rho}$. Contrary to this expectation, the data reveal that $\hat{\rho}$ is frequently negative, as discussed below.

2. Gravitation Toward Evidence or Open Space?

2.1 Data, Source, and Concept

We draw on data from a two-wave quasi-experimental panel study comprising two identical online survey administrations conducted under a unified research design. The combined dataset

includes responses from 1,777 participants—1,073 in Study 1 and 704 in Study 2—allowing results to be interpreted collectively rather than as separate studies (Vlasceanu et al., 2021). Participants were recruited via Cloud Research[3] and asked to evaluate the perceived accuracy of political propositions using a bounded 0–100 scale, where 0 denotes "extremely inaccurate," 100 denotes "extremely accurate," and intermediate values reflect graded accuracy judgments. Each participant rated 36 political propositions, categorized as 12 neutral, 12 Democrat-leaning, and 12 Republican-leaning. To assess changes in evaluative judgment, 36 factual statements were introduced between waves—18 supporting the original proposition (encouraging movement toward the upper bound) and 18 refuting it (encouraging movement toward the lower bound). All participants were exposed to the same factual statement between waves for each proposition they initially evaluated. The sample was divided between Democrats ($N = 904$) and Republicans ($N = 873$), yielding a total of 63,972 observations of change scores ($T^2 - T^1$). This constitutes a repeated-measures pre-test/post-test design, assessing within-subject change between a baseline (pre-evidence) and a follow-up (post-evidence) time point. The study thus adopts a longitudinal framework in which the same latent construct—evaluative judgment—is measured twice using a structurally invariant instrument.

While the original study examined the effects of prediction error on response revision across partisan lines, our analysis focuses on how respondents' initial positions relative to scale boundaries shape the direction and magnitude of subsequent change.

[3]Participants are assumed to have completed the survey independently, without interaction or influence from others.

Table 2: Descriptive Statistics for Study Sample and Design (Vlasceanu *et al.*, 2021*)

Characteristic	**Count / Value Range**
Total Participants (N)	1,777
Total Observations of Change ($N \times 36$)	63,972
Democrats (Party 1)	904
Republicans (Party 2)	873
Total Propositions (12-Dem, 12-Rep, 12 Neutral)	36
Waves	2
Evaluation Scale	[0, 100]
Initial Response Evaluation (pre-evidence)	T^1
Subsequent Response Evaluation (post-evidence)	T^2
Between-Wave Evidence Supporting E_S	18
Between-Wave Evidence Refuting E_R	18
Unit of Analysis	Participant j evaluation of proposition i: (R^1_{ji}, R^2_{ji}) $\in [0, 100] \subset \mathbb{Z}$

† The original survey included additional covariates (e.g. *Supports Trump*, *Belief Resilience*); these are excluded based on feature importance scores derived from an XGBoost machine learning model.

Table 3 differentiates between In-Party Support (IPS) and Out-Party Support (OPS). It uses directional arrows ($\uparrow, \downarrow$) to indicate the direction of observed change, and partisan alignment to denote whether the change supports the respondent's in-party or out-party, given the proposition's leaning. Each change is further classified as evidence-congruent (W) or incongruent (WO), based on the alignment between the direction of change and the polarity of the intervening evidence: W indicates congruent change (e.g., a decrease following refuting evidence), while WO denotes incongruent change (e.g., a decrease following supporting evidence). Each change is classified according to three dimensions: (1) proposition type (Democratic- or Republican-leaning, excluding neutral items), (2) evidence polarity (supporting or refuting), and (3) behavioral response ($\pm\Delta T_{ij}$). Together, these dimensions determine both partisan alignment (IPS or OPS) and evidence congruency (W or WO). Formally, W corresponds to $-\Delta T_{ij} \wedge -E_i$ or $+\Delta T_{ij} \wedge +E_i$, while WO corresponds to $-\Delta T_{ij} \wedge +E_i$ or $+\Delta T_{ij} \wedge -E_i$.

Table 3: Characterized Response Change Relative to Proposition Type, Evidence Polarity, and Political ID

Party ID = Dem.	**Evidence**	**W**	**WO**
Proposition	D	↑ IPS/W	↓ OPS/WO
		↓ OPS/W	↑ IPS/WO
	R	↑ OPS/W	↓ IPS/WO
		↓ IPS/W	↑ OPS/WO

IPS (In-Party Support): A Democrat-affiliated respondent either increases their judgment toward "accurate" for a Democrat-leaning proposition or decreases their judgment toward "inaccurate" for a Republican-leaning proposition. **OPS (Out-Party Support):** A Democrat-affiliated respondent either decreases their judgment toward "inaccurate" for a Democrat-leaning proposition or increases their judgment toward "accurate" for a Republican-leaning proposition. **W (With Evidence):** Change is congruent with evidence-polarity, either decreasing when evidence refutes or increasing when evidence supports. **WO (Without Evidence):** Change is incongruent with evidence-polarity, either decreasing when evidence supports or increasing when evidence refutes. Arrow (↑ or ↓) denotes direction of raw observed change. We exclude neutral propositions, responses that did not change $T_{ji}^{2} = T_{ji}^{1}$, and the normalized change index $\hat{\rho}$ from this exposition.

Each response is classified as either evidence-congruent or evidence-incongruent. Congruent updates follow the direction of the evidence and include: (1) OPS.W—shifts away from in-party positions consistent with the evidence, and (2) IPS.W—shifts toward in-party positions aligned with the evidence. Incongruent updates oppose the evidence and include: (1) OPS.WO—shifts away from in-party positions contrary to the evidence, and (2) IPS.WO—shifts toward in-party positions contrary to the evidence.

Table 4: Observed Response Change Categorization

Main Category	Subcategory	Count
Congruent	IPS.W	13,519
	OPS.W	13,222
Incongruent	IPS.WO	8,805
	OPS.WO	8,102
Total Congruent		**26,741**
Total Incongruent		**15,907**

Excluding neutral propositions—i.e., statements lacking a clear partisan valence as classified by the original researchers—*37% of responses were incongruent with the directional thrust of the intervening evidence.* When neutral items are included, this proportion rises modestly to 38.6%.[4]

The extent of incongruence observed is notable, as it challenges the normative expectation that individuals will, on average, revise their judgments in accordance with credible evidence. Although psychological mechanisms such as motivated reasoning and asymmetric responsiveness to partisan cues may partially account for this counter-directional movement in response to evidence, they do not fully explain the observed pattern. Notably, roughly 48% of incongruent updates do not align with respondents' partisan affiliations, indicating that a substantial portion of such responses cannot be attributed to ideological bias alone.

The prevalence of incongruent updates—particularly those misaligned with partisan interests—suggests that such patterns may not be fully attributable to psychological mechanisms such as motivated reasoning. Instead, these apparent counter-evidential shifts may reflect the asymmetric opportunity space imposed by the response scale, rather than a deliberate rejection of the evidence itself.

[4]Details regarding proposition content and the classification of evidentiary polarity are provided in the Appendix.

2.2 Deliberately Partitioning the Data

To examine position-dependent dynamics, we stratify responses based on whether the initial response T_{ji}^1 lies above or at/below the scale midpoint (50). Specifically, we define two response groups:

$$\mathscr{G}_- = \{(j,i) : T_{ji}^1 \leq 50\}, \quad \mathscr{G}_+ = \{(j,i) : T_{ji}^1 > 50\}$$

Each of the 36 baseline responses from participant j is assigned to one of these groups based on their initial position. We then compute the proportion of responses that increase, decrease, or remain unchanged between waves:

$$\begin{aligned} +\Delta &= \{(j,i) : T_{ji}^2 - T_{ji}^1 > 0\} \qquad \text{(positive change)} \\ -\Delta &= \{(j,i) : T_{ji}^2 - T_{ji}^1 < 0\} \qquad \text{(negative change)} \\ \text{No } \Delta &= \{(j,i) : T_{ji}^2 - T_{ji}^1 = 0\} \qquad \text{(no change)} \end{aligned}$$

We further condition these proportions on the direction of intervening evidence (supportive or refutative).[5] For instance, the proportion of positive updates within the upper group $\mathscr{G}_+$ exposed to supportive evidence ($E_i = E_S$) is given by:

$$P_{+\Delta|E_S,\mathscr{G}_+} = \frac{\#\left\{(j,i) : T_{ji}^1 > 50 \wedge T_{ji}^2 - T_{ji}^1 > 0 \wedge E_i = E_S\right\}}{\begin{array}{c}\#\left\{(j,i) : T_{ji}^1 > 50 \wedge T_{ji}^2 - T_{ji}^1 > 0 \wedge E_i = E_S\right\} + \\ \#\left\{(j,i) : T_{ji}^1 > 50 \wedge T_{ji}^2 - T_{ji}^1 < 0 \wedge E_i = E_S\right\} + \\ \#\left\{(j,i) : T_{ji}^1 > 50 \wedge T_{ji}^2 - T_{ji}^1 = 0 \wedge E_i = E_S\right\}\end{array}}$$

This proportion captures the relative frequency of positive change, conditional on an initial position in the upper half of the scale and exposure to supportive evidence.

Consistent with the criteria outlined in Section 2.1, a response shift is classified as incongruent when its direction contradicts the polarity of the intervening evidence. Table 5 presents the full set of incongruent updates across neutral, Democrat-leaning, and Republican-leaning propositions, disaggregated by initial position: $\mathscr{G}_-$ (at or below the midpoint) and $\mathscr{G}_+$ (above the midpoint).

[5] E_i is a scalar indicating the evidence direction (e.g., $+1$ for supportive, -1 for refutative). Expressions such as $[E_i = E_S]$ denote logical predicates used in conjunction with other conditions via $\wedge$.

Table 5: Two-proportion Z-tests comparing incongruent change by T^1 groups $\mathscr{G}$ and evidence polarity E (E_R - Evidence Refutes and E_S - Evidence Supports)

Test	$\mathscr{G}_-$	$\mathscr{G}_+$	**95% CI**	χ^2**, df,** p
$+\Delta \wedge E_R$	0.428	0.202	[0.216, 0.236]	1897.9, 1, $p < .001$
$-\Delta \wedge E_S$	0.159	0.409	$[-0.260, -0.241]$	2117.8, 1, $p < .001$

Prop 1: $T^1 \leq 50 \equiv \mathscr{G}_-$; Prop 2: $T^1 > 50 \equiv \mathscr{G}_+$. Confidence intervals reflect differences in proportions (Prop 1 − Prop 2).

For $\mathscr{G}_-$ within the context of incongruent updates ($+\Delta \wedge E_R$ or $-\Delta \wedge E_S$), the proportion of updates *away from the nearest bound* (i.e., lower bound) over all incongruent changes is $\frac{.428}{.428+.159} \approx .73$. For $\mathscr{G}_+$, incongruent updates ($+\Delta \wedge E_R$ or $-\Delta \wedge E_S$), the proportion of updates *away from the nearest bound* (i.e., upper bound) over all incongruent changes is $\frac{.409}{.409+.202} \approx .67$. The data shows that in both groups, *misinterpretation is more likely when change moves away from the nearest bound—toward "more open values."*

Table 6: Full observed change proportions by T^1 group and evidence type, including the *No Change* ($T^1_{ji} = T^2_{ji}$) category

Group	$+\Delta \wedge E_R$	$-\Delta \wedge E_R$	No $\Delta \wedge E_R$	$+\Delta \wedge E_S$	$-\Delta \wedge E_S$	No $\Delta \wedge E_S$
$\mathscr{G}_-$	0.428	0.488	0.084	0.799	0.159	0.042
$\mathscr{G}_+$	0.202	0.727	0.071	0.482	0.409	0.109

Proportions within each row sum to 1 for each evidence condition (E_R: Evidence Refutes; E_S: Evidence Supports). E_R: $\chi^2 = 2055.974$, df = 2, $p < .001$. E_S: $\chi^2 = 3057.714$, df = 2, $p < .001$.

Notably, evidence-congruent response shifts are more likely to occur in the direction away from the nearest scale boundary as well. Among cases in $\mathscr{G}_-$, approximately 62% of congruent updates ($\frac{0.799}{0.799+0.488} \approx 0.62$) move upward—away from the lower bound. Likewise, in $\mathscr{G}_+$, about 60% of congruent updates ($\frac{0.727}{0.727+0.482} \approx 0.60$) move downward—away from the upper bound. In both

subgroups, the majority of evidence-congruent and incongruent response shifts are directed away from the nearest scale boundary, revealing a consistent directional asymmetry.

While the observed asymmetry is clear, its source remains uncertain, though it is distinct from classical regression to the mean treatment.

As we will demonstrate throughout, the observed asymmetries do not necessarily arise from stochastic processes, but rather from the structural constraints—or differential affordances for change—imposed by a respondent's initial proximity to the scale boundaries. This asymmetry is structural rather than stochastic in origin. For this section, the conventional assumption of a universal central tendency serves as the central basis for our departure from standard RTM interpretations.

Responses are now grouped by party affiliation (Democrat or Republican) and by evidence type (supportive or refutative). As we will observe, the expected reversion point is not fixed, but varies systematically across subgroups and baseline response positions. To investigate this relationship, we apply LOESS (Locally Estimated Scatterplot Smoothing) to model the association between initial response values (T^1) and the average observed change (ΔT) within each subgroup.

LOESS is a nonparametric regression technique that fits low-degree polynomials to localized subsets of the data using weighted least squares. This approach captures local structure in the relationship between variables without imposing a global parametric form. Its flexibility enables the detection of nonlinear trends and potential inflection points in the trajectory of response change as a function of baseline position.[6]

Formally, let $T^1 \in \mathbb{Z}$ denote the initial response, and let $\Delta T = T^2 - T^1$ represent the observed change. Let $\mathscr{G}$ denote the set of subgroups defined by the cross-classification of party affiliation and evidence type. For each subgroup $g \in \mathscr{G}$, we estimate the smoothed function:

$$\hat{f}_g(T^1) = \text{LOESS}(\Delta T \mid T^1; g)$$

[6]LOESS can operate over discrete predictors (e.g., integers), but conceptually treats the predictor as continuous to estimate local trends.

Here, $\hat{f}_g(T^1)$ represents the locally estimated mean change conditional on baseline response T^1 within subgroup g. This function captures how the direction and magnitude of change vary across the response scale for each subgroup, without assuming a parametric relationship.

Our analytic goal is to identify points along the scale where the expected change reverses direction—that is, values T^1 where the LOESS curve crosses zero. Formally, we posit that,

$$\exists T^{1*} \in [a,b] \text{ such that } \hat{f}_g(T^{1*}) = 0 \quad \text{and} \quad \frac{d}{dT^1}\hat{f}_g(T^{1*}) \neq 0.$$

Such points T^{1*} indicate nonstationary reversals in the direction of expected change, conditional on subgroup membership. The nonzero derivative ensures that the function crosses zero—indicating a true directional reversal—rather than merely touching it.

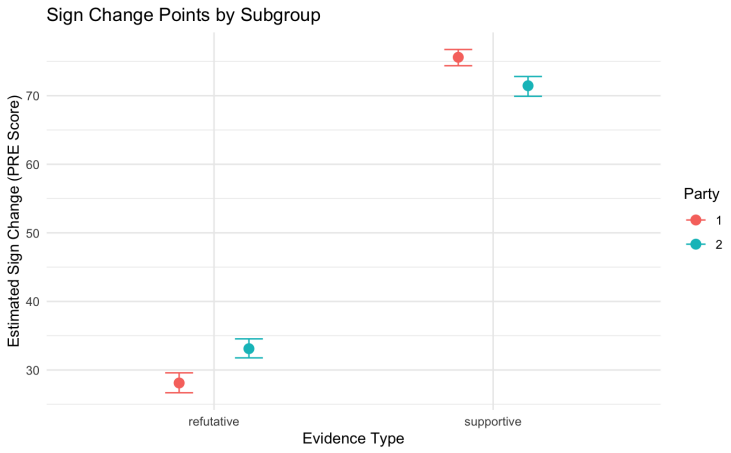


Figure 5: Mean inflection points in response change ($\pm\Delta T$) are plotted as a function of initial response values (T^1), conditional on party affiliation (indicated by color) and evidence type (separated along the x-axis). Party 1 (orange) corresponds to Democrats, and Party 2 (cyan) corresponds to Republicans.

Table 7 presents the corresponding results summarized in Figure 5.

Party	Evidence Type	T^{1*} Point Estimate	CI Lower	CI Upper	CI Width
1	Refutative	28.11	26.68	29.59	2.91
2	Refutative	33.11	31.77	34.54	2.77
1	Supportive	75.60	74.37	76.74	2.37
2	Supportive	71.45	69.91	72.80	2.89

Table 7: Bootstrapped estimates of sign-change points with 95% confidence intervals, disaggregated by party affiliation and evidence type. All participants self-identified as either Democrats (Party 1) or Republicans (Party 2); no third-party or unaffiliated respondents are present in the dataset.

The initial response value T^1 at which the average change (ΔT) reverses sign varies systematically across groups, indicating that response trajectories depend jointly on party affiliation, evidence polarity, and baseline position. These reversal points are non-uniform, reflecting heterogeneous and context-dependent thresholds for directional updating.

The bootstrapped 95% confidence intervals ($\alpha = 0.025, 0.975$) for the estimated sign-change points exclude both the scale midpoint (50) and the population mean (≈ 52.5), providing evidence against the hypothesis of a central reversion point.[7] Additionally, a Lord–Novick (1968) excess-change test reveals an average excess change of -2.82 scale points. While modest in magnitude on a 0–100 scale, this deviation is statistically significant ($t = -22.12$, $p < .00001$), indicating systematic departures from classical regression-to-the-mean expectations. These findings challenge the applicability of standard RTM models in this context (see Appendix, Table 32).

The descriptive similarity in sign-reversal patterns across partisan groups suggests that the observed inflection points may not primarily reflect differences in information processing, but rather structural or contextual features common across groups. Under supportive evidence, average response trajectories reverse from increasing to decreasing near the upper boundary of the third

[7] Confidence intervals for each party-by-evidence condition were computed using 1,000 nonparametric bootstrap resamples. Each interval reflects the central 95% of the empirical distribution, with 2.5% of estimates falling below the lower bound and 2.5% above the upper bound.

quartile (Q3) for both partisan groups,[8] indicating a downward "preemptive" adjustment even in the presence of supportive evidence. Conversely, under refutative evidence, reversals from decreasing to increasing emerge near the lower end of the second quartile (Q2), reflecting upward deviation despite evidence favoring continued decline. In both cases, the inflection points appear to reflect sensitivity to the proximity of scale boundaries rather than partisan alignment or deliberate rejection of evidence.

Crucially, as demonstrated, the observed pattern does not imply universal convergence toward the population mean. While less extreme responses may follow more extreme ones, the asymmetries in response change appear to be driven primarily by respondents' proximity to the scale boundaries. Near these limits, directional change is structurally constrained, yielding systematic, position-dependent biases. Such boundary effects constitute geometric artifacts of the response scale's bounded structure, rather than random fluctuations around a central tendency as assumed under classical RTM frameworks.

For clarity, we provide formal definitions of both regression to the mean and open movement bias.

- Classical Regression to the Mean (RTM): A statistical phenomenon that occurs when repeated measurements are imperfectly correlated, such that extreme values on an initial measurement are, on average, followed by less extreme values on subsequent measurements. This tendency arises not from systematic change, but from random error, measurement noise, or natural variability in the construct being measured (Stigler, 1997).

- Boundary-Induced Asymmetry (Open Movement Bias): A structural effect of bounded response scales, whereby proximity to a scale endpoint constrains the feasible range of movement, systematically biasing responses away from the nearest boundary and toward the direction of greater available scale space, or "open support." This results in non-random, directionally asymmetric shifts that are independent of measurement error, central tendency,

[8] Inflection points were estimated using second-derivative approximations via finite differences within a ±5-point window.

or intervening evidence. Non-random, directionally asymmetric refers to shifts that are systematically biased in one direction—such as upward or downward—based on the respondent's initial position on the scale, rather than being symmetrically distributed around zero or the midpoint.

In survey research, response patterns are rarely acknowledged as outcomes of structural properties intrinsic to the measurement space. However, the affordances and constraints imposed by scale geometry can systematically influence the distribution and direction of observed response change. To help disentangle substantive change from scale-induced distortion, we normalize each respondent's shift by the maximum feasible movement from their initial position.

2.3 Introducing the Diagnostic $\hat{\rho}$

To analyze response dynamics relative to the scale's constraints, we group responses into bins based on their baseline position. For each bin, we compute the mean observed change and normalize it by the maximum possible change from that bin's center.

The response scale is the discrete bounded interval $[0,100] \subset \mathbb{Z}$, and each response is indexed as

$$T_{ji}^{t} \in [0,100], \qquad j \in \mathscr{J}, \quad t \in \{1,2\},$$

where T_{ji}^{t} denotes the response of individual j to proposition i at wave t. The set $\mathscr{J}$ indexes all individuals.

The within-person change in response is defined as:

$$\Delta T_{ji} = T_{ji}^{2} - T_{ji}^{1}.$$

We define bins centered at points $b_0 \in [0,100]$, with a fixed half-width $c > 0$. Each bin includes all respondent-item pairs whose initial response falls within the interval $[b_0 - c,\ b_0 + c]$. Formally,

we define,

$$\text{bin}(b_0) = \left\{(j,i) \in \mathscr{J} \times I \,:\, T_{ji}^1 \in [b_0 - c,\, b_0 + c]\right\}.$$

This construction allows us to summarize response-change behavior across all propositions $i \in I$, conditional on initial response level. For each bin, we calculate the average change $\widehat{\Delta T}$, and normalize it by the maximum possible shift from b_0 to the furthest endpoint of the response scale.

In practice, we set the bin center spacing and half-width to $c = 2.5$ partitioning the response scale into 20 equal-width bins of 5 units each. The centers of these bins are:

$$b_0 \in \{2.5,\, 7.5,\, 12.5,\, \ldots,\, 97.5\}.$$

The size of the bin, the number of responses falling within the interval centered at b_0 (i.e., the cardinality of the bin), is given by:

$$n_{b_0} = |\,\text{bin}(b_0)|$$

We drop 'bin' from $\text{bin}(b_0)$ for notational convenience. Herein we allow b_0 to refer to both the bin center and the subinterval centered at that value.

The average of individual differences within each bin is:

$$\widehat{\Delta T}(b_0) = \frac{1}{n_{b_0}} \sum_{T_{ji}^1 \in b_0} \left(T_{ji}^2 - T_{ji}^1\right). \tag{4}$$

where $\widehat{\Delta T}(b_0)$ denotes the mean observed change for all responses originating from the bin centered at b_0.

To allow unbiased comparisons in the magnitude of response change across bins, we define the bin-normalized change index (a reinterpretation of Equation 1) as follows:

$$\hat{\rho}(b_0) = \frac{\widehat{\Delta T}(b_0)}{\max\big(100 - b_0,\, b_0 - 0\big)} \in [-1, 1]. \tag{5}$$

where 0 and 100 represent the lower and upper bounds of the response scale. The normalization adjusts for the fact that the maximum possible change from any starting point is constrained by its distance to the more distant boundary. The index $\hat{\rho}(b_0)$ captures the average change within each bin as a proportion of its feasible range, yielding a scale-relative measure bounded in $[-1, 1]$. This enables a fair, scale-relative comparison of binned average response changes across the full range of initial positions by adjusting for the structural constraints on feasible movement.

By construction, $\hat{\rho}(b_0)$ captures both the average direction, via its sign, and the average magnitude of change originating from each bin. By aggregating responses into bins and estimating average change normalized by the structurally constrained potential for movement, $\hat{\rho}(b_0)$ serves as a nonparametric plug-in estimator of position-conditional updating. Because the estimator operates on bin-level aggregates rather than individual trajectories, it avoids complications associated with individual-level coupling or idiosyncratic noise. Note, *the $\hat{\rho}_{ij}$ index is computed throughout the exposition using a leave-one-out (LOO) procedure, ensuring that the dependent variable is never reintroduced into the set of regressors.*[9] This approach necessarily sacrifices individual-level variance in favor of capturing localized patterns.

This approach bears additional interpretive caveats. A low value of $\hat{\rho}(b_0)$ may reflect either (i) consistent movement toward the nearest bound with modest average change, or (ii) heterogeneous shifts in both directions that cancel out when averaged across n_{b_0}. Care is therefore required in interpretation. In general, bins with larger sample sizes and narrower internal variance support more reliable inference.

When combined with between-wave evidence, treatments, or covariates, the normalized change index functions as a diagnostic tool. By scaling observed shifts relative to the maximum feasible change from each bin's initial position, it adjusts for unequal movement potential across the scale and clarifies how initial responses shape the distribution of change, thereby enhancing the detection of localized response patterns.

[9] While this approach avoids direct coupling of individual pre- and post-scores, it still conditions on initial value bins, which may introduce selection bias if those bins correlate with unobserved variables.

In a random system, changes would be symmetrically distributed around zero, indicating no directional bias. While normalization improves comparability in the extent of change across bins, it does not eliminate the directional asymmetry imposed by the bounded scale. The index adjusts for variation in change magnitude but not for the unequal distribution of reachable outcomes. This *directional asymmetry* is structurally embedded in the response geometry and arises prior to any second-wave measurement—it cannot be fully corrected post hoc.

Assumptions

Response behavior can exhibit asymmetries in ”true” change driven by factors such as ideology or motivated reasoning. Consequently, $\hat{\rho}$ may overcorrect for structural constraints or mistakenly attribute genuine psychological asymmetries to geometric affordances.

Equations [4] and [5] assume no systematic dropout between waves $t = 1$ and $t = 2$. We assume and assign equal width bins. The use of max in [5] ensures a conservative estimate of proportional change by referencing the more distant boundary.

2.4 Implementing $\hat{\rho}$

The response scale is partitioned into 20 bins, each spanning a 5-point interval and centered on its respective midpoint. When unaffected by exclusions, each bin contains over 1,600 observations. Figure 6 organizes results by party affiliation (rows) and by evidence condition (supportive vs. refutative, shown in separate panels). Summary statistics for all 80 bin combinations are provided in Appendix Table 21. Tile colors indicate mean response change from Wave 1 to Wave 2 $(\bar{\Delta}T)$: green signifies an increase, orange a decrease, and white a negligible shift. Overlaid arrows visualize the normalized change index $(\hat{\rho})$; arrow direction reflects the sign of change, and length indicates relative magnitude.

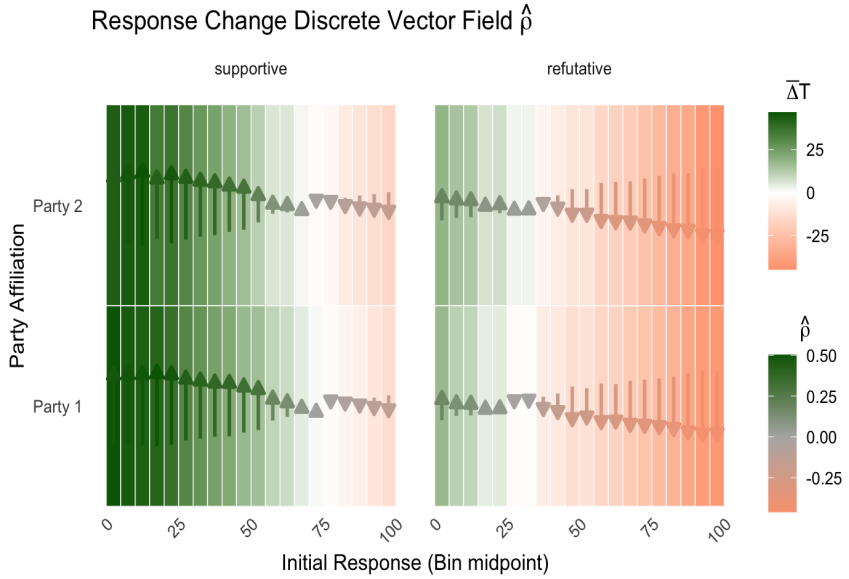


Figure 6: Descriptive bin-level statistics for mean raw change $\bar{\Delta}T$ and normalized change $\hat{\rho}$ are presented. The two measures exhibit a strong positive correlation at the bin level ($r \approx 0.98$), as illustrated in the correlation matrix in Appendix Figure 26.

A clear pattern emerges: bins with lower initial values show positive average changes, while higher-value bins show negative average changes—revealing a monotonic gradient in response shifts structured by initial position. Though the effect size varies by evidence type, bin, and partisan affiliation, the overall trend holds across the response range. Figure 6 highlights three central dynamics: (i) the direction of average change is systematically associated with initial bin location—positive shifts are more common in lower bins, while negative shifts predominate in higher bins; (ii) incongruent response patterns tend to emerge in the lower quartile under refutative evidence and in the upper quartile under supportive evidence; and (iii) these patterns are fairly consistent across partisan groups.

To assess how initial position on the response scale moderates the magnitude and direction of response change, we aggregate individual responses into quartile-based bins according to their baseline values. This coarser grouping enables tractable comparisons of normalized change across larger segments of the response distribution while preserving information about proximity to the scale's endpoints.

We stratify these quartile-level aggregates by party affiliation and evidence condition, allowing us to evaluate whether response dynamics differ systematically by initial position, political identity, or the polarity of intervening information. Figure 7 visualizes the normalized change index $\hat{\rho}$ as a function of baseline quartile, disaggregated by party and evidence type.

To complement these visual patterns, we conduct pairwise comparisons of mean absolute change ($|\Delta T|$) across quartiles using Tukey's Honest Significant Difference (HSD) procedure. The results, presented in Table 8, provide inferential support for systematic differences in responsiveness by initial position.

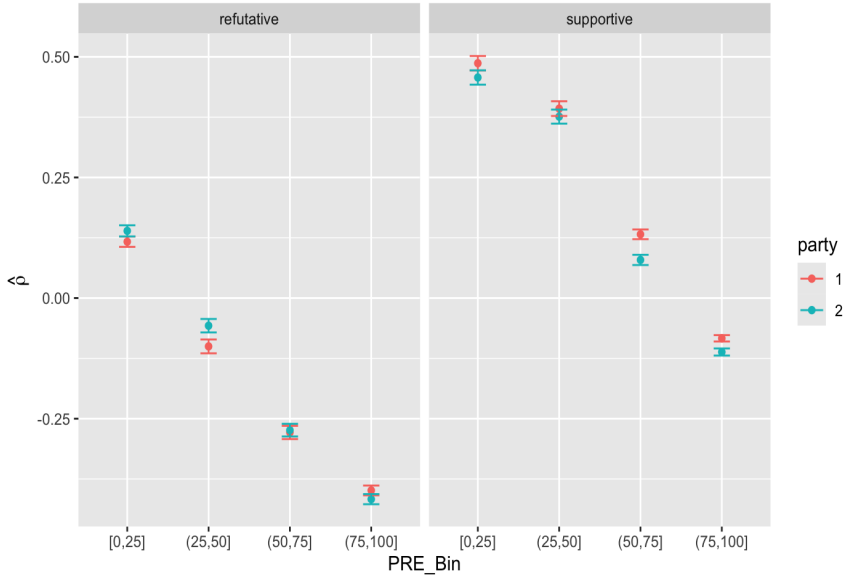


Figure 7: Point color indicates partisan affiliation (Party 1 = Democrat; Party 2 = Republican), while panel columns distinguish evidence conditions—supportive (left) and refutative (right). Corresponding descriptive statistics are provided in Appendix Table 31.

Table 8: Tukey HSD Pairwise Comparisons of Mean Absolute Change by Baseline Quartile

Comparison	**Difference**	**95% CI**	**Adj. p-value**
2 vs. 1	-5.07	$[-5.78, -4.37]$	$<0.0001^{***}$
3 vs. 1	-6.11	$[-6.80, -5.41]$	$<0.0001^{***}$
4 vs. 1	-0.45	$[-1.16, 0.26]$	0.368
3 vs. 2	-1.03	$[-1.74, -0.33]$	0.0009^{***}
4 vs. 2	4.62	[3.91, 5.34]	$<0.0001^{***}$
4 vs. 3	5.66	[4.95, 6.37]	$<0.0001^{***}$

Note. Results reflect pairwise comparisons of mean absolute change scores ($|\Delta T|$) across baseline quartiles. Significance codes: $^{***}p < 0.001$.

The results support the hypothesis that the effects of evidence are contingent on respondents' initial positions on the response scale. Absolute change varies systematically by baseline quartile: individuals in the extreme quartiles (Q1 and Q4) exhibit greater average movement than those in the middle quartiles (Q2 and Q3).

Interestingly, *mean responses in Q1 and Q4 consistently shift away from the nearest scale boundary, irrespective of evidence condition.* While both quartiles demonstrate greater responsiveness than Q2 and Q3, the absence of a statistically significant difference in absolute change between Q1 and Q4 suggests an approximately symmetric boundary effect.

A key limitation of this approach lies in disentangling the structural tendency of the first and fourth quartiles to exhibit directional movement away from the nearest scale boundary. Specifically, reliance on the mean renders the analysis sensitive to outliers: a single large shift can disproportionately affect the average, masking the prevalence of smaller but more frequent movements in the opposite direction. For example, one respondent shifting from 1 to 30 can offset four respondents each shifting from 5 to 0, resulting in a mean change of $\bar{\Delta}T = 2$.

We address this limitation via binomial tests (Table 30 in Appendix), which reveal that respondents in the extreme quartiles—Q1 and Q4—show *directional shifts* toward the opposite bound at rates significantly above chance: 64.0% of Q1 responses shift upward, and 66.8% of Q4 shift downward. These results support the hypothesis that low initial scores tend to increase, while high initial scores tend to decline from wave 1 to 2, independent of the evidence polarity ($\pm$).

Even after normalizing for scale-constrained movement, results suggest residual structural asymmetries in response change. Despite accounting for each respondent's maximum feasible movement (MFM), directional biases persist. Proximity to scale boundaries appears to influence not only the opportunity for movement, but also the likelihood, direction, and magnitude of observed change.

Returning to a more granular level of analysis, Figure 8 displays bootstrapped means and 95% confidence intervals for the normalized deviation index, disaggregated by 20 initial bin subintervals and party affiliation.

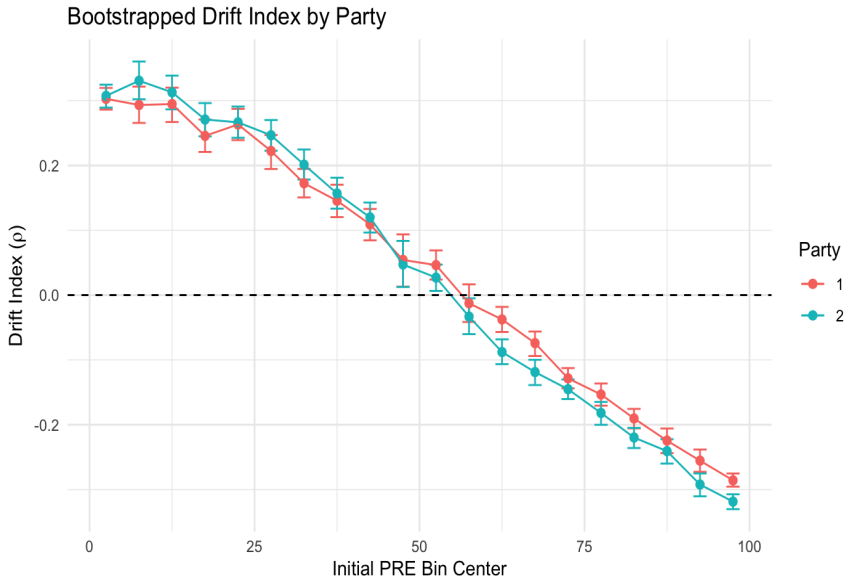


Figure 8: Normalized deviation across 20 discrete bins separated by party. For a 20 bin summary corresponding to Figure 8 see Appendix Table 29.

Lower values of the normalized change metric, $\hat{\rho}$, are expected near the scale boundaries due to the larger maximum feasible movement (i.e., a larger denominator). Yet, contrary to this

expectation, respondents near the extremes tend to exhibit greater proportional shifts than those near the center—even though achieving the same normalized value requires a larger absolute change. For example, a shift from 0 to 20 produces $\hat{\rho} = .2$, just as a shift from 50 to 60 does, despite the former involving a greater raw change. Empirically, the absolute value of $\hat{\rho}$ tends to decline as initial responses approach the scale midpoint, indicating an attenuation in responsiveness that is counterintuitive given the metric's symmetrical structure. This pattern suggests that proximity to scale boundaries may amplify responsiveness at the bin level. This does not preclude the presence of floor or ceiling effects, as bin-level averages may obscure a substantial proportion of static responses clustered near the scale's boundaries. Nonetheless, the elevated mean responsiveness observed in boundary-adjacent bins is noteworthy. Although some respondents may exhibit no change, those who do update tend to do so with relatively large magnitudes—consistent with the directional affordances of the bounded response scale. Moreover, these averages are likely attenuated by the inclusion of inert responses, which suppress the estimated magnitude of change.

Bins near the scale's endpoints exhibit greater responsiveness—both in absolute and normalized terms—than central bins. Consistent with affordance theory, the response scale's structure appears to influence behavior, offering position-dependent opportunities for movement. Figures 7 and 8, along with binomial test results, point to an "open-movement bias": respondents' updates reflect sensitivity to structural affordances inherent in their starting position, rather than being solely driven by individual characteristics or external inputs.

2.5 Reversibility

To assess the structural dependence of response change on initial values—i.e., the potential "irreversibility" of starting positions—we conduct a non-parametric permutation test of structured variance. This test evaluates whether observed variation in mean change and normalized deviation $\hat{\rho}$ across initial bins reflects genuine dependence on baseline responses (T^1), or could plausibly arise from random variation.

Observed variance in mean raw change across baseline bins was substantial ($\text{Var}_{\text{obs}} = 333.83$), as

was the variance in mean normalized change ($\text{Var}_{\hat{\rho}} = 1.5745$), indicating considerable heterogeneity linked to initial response position. To assess whether such structure could arise by chance, we conducted a permutation test (5,000 iterations) under the null hypothesis that change scores are randomly distributed across bins. The resulting null distribution was tightly centered near zero, with no simulated variances approaching the observed values. Across all bins and both party groups, permutation p-values were < 0.0002, strongly rejecting the null. These results reinforce the claim that baseline position systematically shapes response change by imposing geometric constraints on movement.

In sum, asymmetries in response change are both empirically robust and structurally patterned, supporting three central claims. First, variance in raw and normalized change is maximized under the observed (non-permuted) bin structure, indicating systematic dependence between initial position and subsequent change. Second, the persistence of asymmetry even after normalization by maximum feasible movement (MFM) suggests that these patterns are not reducible to boundary constraints alone but reflect a deeper structural interaction between the measurement instrument and behavioral response processes. Third, the consistency of effects across bins and party affiliations—paired with their improbability under permutation-based null models—rules out random error or sampling noise as primary drivers. Together, these findings underscore the role of geometric affordances in shaping response dynamics.

Having examined ex post response patterns, we now turn to whether the structure of the response scale *itself* systematically shapes (i.e., interacts with) *the process of change.* While latent class factor analysis (LCFA) and structural equation modeling (SEM) are widely used to model latent traits and response behavior, they typically assume symmetric and continuous response processes. These methods do not natively accommodate directional, position-dependent shifts in ordinal change scores that emerge from the structural features of bounded response scales. This conceptual and methodological limitation motivates the development of the PLRC framework.

2.6 Piecewise Linear Regression Corridor

We estimate a piecewise linear regression model with threshold points at L and U, permitting distinct slopes in the segments below and above these points. Although piecewise linear models with threshold effects are well-established in statistical modeling, their application to survey response dynamics, particularly in capturing position-dependent change, is comparatively underexplored.

This model predicts change in response ($\Delta T_{ji} = T_{ji}^2 - T_{ji}^1$) based on the baseline score T_{ji}^1 and its position relative to two interior points L and U on the scale such that $LB < L < U < UB$, where LB and UB represent the lower and upper bounds of the response scale, respectively. We select the threshold points L and U based on the empirical inflection points identified in Figure 5 and Table 7, and compare model performance across alternative thresholds to assess the graded nature of response change across the scale. While our approach uses theory-guided selection, these thresholds could alternatively be estimated non-parametrically or through data-driven optimization procedures. These thresholds define a neutral corridor on the bounded response scale $[L,U]$, within which change is not attributed to structural bias. Deviations observed outside this corridor are interpreted as partially structurally induced, reflecting both respondent-level dynamics and the asymmetric constraints imposed by proximity to the boundaries.

The formula implemented is:

$$\Delta T_{ji} = \begin{cases} \alpha + \beta_L (L - T_{ji}^1) + \varepsilon_j & \text{if } T_{ji}^1 < L \\ \alpha + \varepsilon_j & \text{if } L \leq T_{ji}^1 \leq U \quad \text{for } T_{ji}^1 \in [LB, UB] \\ \alpha + \beta_U (T_{ji}^1 - U) + \varepsilon_j & \text{if } T_{ji}^1 > U \end{cases} \tag{3}$$

where $\Delta T_{ji} = T_{ji}^2 - T_{ji}^1$ is the observed change in response for individual j on topic i, T_{ji}^1 is the baseline response, L and U are the inner corridor thresholds, β_L captures the influence of deviation when $T_{ji}^1 < L$, β_U captures the influence of deviation when $T_{ji}^1 > U$, α is the intercept term, $\varepsilon_j \sim N(0, \sigma^2)$ is the residual error term.

When $T_{ji}^{1} \in [L,U]$, both of the conditional deviation terms drop out, and the model simplifies to: $\Delta T_{ji} = \alpha + \varepsilon_j$. In this case, no deviation penalty is applied and change ΔT_{ji} is modeled as pure noise around a constant α. This specification assumes that individuals whose baseline values fall within the corridor $[L,U]$ exhibit a constant expected change—i.e., they represent structurally stable cases. This permits targeted modeling of deviation only in the tails of the baseline distribution, where structural asymmetries appear most pronounced.

As is often the case, there is a tradeoff between interpretability and model flexibility, governed by the choice of the threshold interval $[L,U]$. A wider corridor results in more observations falling within the interval, where the model assigns a constant expected change $\Delta T_{ji} = \alpha$. This reduces the model's responsiveness to variation in T_{ji}^{1} across its full range, but increases the stability and precision of the baseline estimate α. In contrast, a narrower corridor increases the number of observations that activate the deviation terms β_L and β_U, allowing the model to capture more nuanced changes in ΔT_{ji} for extreme values of T_{ji}^{1}. However, this can reduce the precision of those slope estimates, especially when the number of observations in the tails is small. In the limiting case where only one observation lies below L and one above U, the model captures the maximum possible deviation effects at the extremes of T_{ji}^{1}, but offers minimal ability to generalize beyond those boundary points.

However, the objective of this analysis is not to optimize predictive performance. Rather, the model is employed as a tool to illustrate how the functional form of predicted change is contingent upon the specification of threshold parameters L and U. Varying these thresholds systematically alters both the assignment of deviation classifications and the interpretation of change magnitudes, thereby demonstrating the model-dependent and initial value (sub-interval) specific nature of inference.

Table 9: Comparison of PLRC Models with Varying Thresholds using Study 1, Vlasceanu, et al.

Model	**Parameter**	**L25_U75**	**L40_U60**	**L10_U90**	**L5_U95**
Correlation	Pearson's r	1.000	0.945	0.878	0.807
			1.000	0.772	0.703
				1.000	0.973
					1.000
Deviation Label Counts	Total deviation labels = 0		12,813		
	Total deviation labels = 1		19,907		
	Total deviation labels = 2		15,846		
	Total deviation labels = 3		3,996		
	Total deviation labels = 4		11,410		
Slope Estimates	β_L	1.6848	0.8194	4.3520	5.4102
	p_L	0.0000	0.0000	0.0000	0.0000
	β_U	-4.9010	-3.4518	-6.2100	-6.3760
	p_U	0.0000	0.0000	0.0000	0.0000

As baseline values T^1 diverge from the corridor threshold points L and U—especially under narrower corridor definitions (e.g., $[40, 60]$)—the distance between T^1 and the nearest threshold increases. Consequently, smaller slope coefficients (β) can explain a given amount of observed change. Conversely, under wider corridors (e.g., $[5, 95]$), the reduced discrepancy requires larger coefficients to capture equivalent variation.

This inverse relationship between threshold proximity and slope magnitude arises from the geometry of the scale. Variation in slope estimates across corridor widths is not necessarily evidence of behavioral change, but rather a structural identification artifact—arising from the role of threshold placement in rescaling the predictor.

Crucially, this geometric artifact does not undermine the model's substantive insight. The

directional asymmetry in slope signs remains meaningful: when $T^1 > U$, the upper-bound term $\beta_U(T^1 - U)$ drives downward predictions, aligning with observed decreases at the high end of the scale; when $T^1 < L$, the lower-bound term $\beta_L(L - T^1)$ contributes positively, consistent with upward shifts from low starting points.

Importantly, the magnitude of β_U consistently exceeds that of β_L across all corridor specifications, despite their symmetric positioning. This asymmetry suggests that responses exceeding the upper threshold are more strongly corrected downward than lower-threshold exceedances are adjusted upward—revealing a systematic, directional bias.[10]

To further characterize these dynamics, we examine deviation label counts—that is, the number of models (out of four) in which a given observation is classified as deviating from the corridor. An observation is labeled as "deviating" within a given model if its baseline value T^1 satisfies $T^1 < L$ or $T^1 > U$. Because each model defines its own corridor, a single observation may be labeled as deviating in multiple configurations.

For example, an observation might lie outside the intervals defined by $[25, 75]$, $[40, 60]$, and $[10, 90]$, but still lie within $[5, 95]$. In this case, it would be labeled as deviating by three of the four models. As the geometry of the thresholds changes—for instance, comparing $[25, 75]$ to $[5, 95]$—the agreement among models decreases. This is reflected in the declining correlation between predicted deviation magnitudes, such as a drop from 1.0 to 0.807 between the L25_U75 and L5_U95 models. Thus, while the predictions are highly correlated overall, they are not interchangeable.

Under the widest corridor specification ($[5, 95]$), these cases exhibit small discrepancy values ($L - T^1$ or $T^1 - U$), and are associated with relatively large slope estimates. This pattern reflects the model's need to account for meaningful observed change originating from those extremes. Notably, 11,410 out of 63,972 observations (approximately 17.8%) fall into this category, indicating that a nontrivial portion of the sample begins in positions where structural constraints are likely to shape

[10]The intercept parameter α is included in each model and serves to shift the predicted response surface. However, α does not influence the classification of observations as deviationing or not; deviation labels are determined solely by the position of the baseline T_i^1 relative to the model's $[L, U]$ interval.

the magnitude and direction of change.

To investigate this dynamic, we examine the proportion of incongruent response shifts at varying threshold values of the baseline measure T^1 likelihood of directionally incongruent change varies as a function of initial position.

Table 10: Proportion of Faulty Changes by Threshold Baseline Values, from Vlasceanu et al.

Threshold	**Condition**	**Proportion of Faulty Changes**
Upward Change under Refutative Evidence		
$T^1 < 5$	Refutative	0.539
$T^1 < 10$	Refutative	0.557
$T^1 < 25$	Refutative	0.525
$T^1 < 40$	Refutative	0.466
Downward Change under Supportive Evidence		
$T^1 > 95$	Supportive	0.510
$T^1 > 90$	Supportive	0.531
$T^1 > 75$	Supportive	0.506
$T^1 > 60$	Supportive	0.445

These results reveal a pattern: the proportion of incongruent, or "faulty," changes increases as baseline responses approach the bounds of the scale. Under refutative evidence, upward shifts exceed 50% for responses below 25, peaking at 55.7% for $T^1 < 10$. Similarly, under supportive evidence, downward shifts surpass 50% for responses above 90. This directional reversal—counter to the implied movement of the evidence—suggests a boundary-induced bias. As responses near a scale endpoint, the likelihood of shifting away from that boundary increases, consistent with structural constraints embedded in the response format.

Notably, as the corridor narrows, the proportion of incongruent changes generally declines,

perhaps because the permissible range for evidence-consistent change increases.

Taken together, these results do not establish a causal effect of geometry on behavior. However, they do suggest that response scale structure—particularly proximity to its limits—plays a measurable role in conditioning the observed relationship between baseline values and subsequent change. In this context, classifications of slope magnitudes and directions, as well as deviations from evidence, may reflect both behavioral variability and the influence of structural affordances embedded in the response format.

We next benchmark the PLRC against four models via simulation.

2.7 Benchmarking PLRC vs 4 Alternatives

The PLRC model is well-suited for estimating change in ordinal or discretized survey responses for several key reasons. First, it captures asymmetry in response dynamics by assigning separate slope terms to upper and lower baseline values, accommodating the fact that change from low versus high initial positions often differs in magnitude and direction. Second, PLRC does not assume continuity, making it appropriate for ordinal data such as Likert scales, where discrete and bounded values may violate assumptions of smooth functional forms. Instead, it defines a stable zone $[L, U]$ in which no structurally induced change is expected, and models deviation only outside that corridor. Third, the model naturally handles ceiling and floor effects by incorporating directional slope terms for baseline responses that fall outside the "stable" range. Instead of treating ceiling/floor effects as residual artifacts or censoring problems, PLRC models them directly as structurally distinct regimes. Finally, PLRC's structure is both simple and interpretable.

To assess model performance, we conduct a Monte Carlo benchmarking study using synthetically generated panel data designed to emulate ordinal survey responses. In each simulation, latent baseline scores $L_1 \sim N(0,1)$ are independently and identically distributed (i.i.d.), and individual-level change scores $\delta \sim N(\mu_d, \sigma_d^2)$ are also drawn i.i.d., where μ_d controls the average true change and σ_d captures inter-individual variability. Post-change scores are computed as $L_2 = L_1 + \delta$ and

transformed via a logistic function to the intervals $[0, 100]$ and $[1, 7]$, respectively, approximating VAS and Likert-type scales. Gaussian noise $\varepsilon \sim N(0, \sigma^2_{\text{noise}})$ is added i.i.d. to each transformed value before discretizing the results into 100- or 7-point ordinal responses using equal-width binning. The resulting outcomes, T^1 and T^2, preserve a known latent structure, enabling evaluation of model accuracy using bias and root mean squared error (RMSE) relative to the true change scores across repeated simulations.

Each simulated dataset is analyzed using a range of change estimation models, including ordinary least squares (OLS), piecewise linear regression with thresholds (PLRT), regression to the mean (RTM), latent change score modeling (LCS) via structural equation modeling, and beta regression. For each simulation, the target change parameter is estimated and evaluated in terms of bias and RMSE. Repeating this process over 2,000 independent simulations provides a basis for assessing model performance. Results are aggregated to benchmark the relative performance of each method across these evaluation criteria.

Table 11: Comparison of Model Outputs on Bias and RMSE Estimates (Scale Range: [1, 7])

Model	**Bias**	**RMSE**	**Bias SD**	**RMSE SD**
Beta	-0.4999	0.6550	0.0218	0.0189
LCS	-0.4999	1.1266	0.0218	0.0295
OLS	-0.4999	1.1266	0.0218	0.0295
PLRT	0.1847	0.6874	0.0491	0.0332
RTM	-0.4999	1.1266	0.0218	0.0295

Table 12: Model Comparison: Bias and RMSE Estimates (Scale Range: [0, 100])

Model	Bias	RMSE	Bias SD	RMSE SD
Beta	-0.4999	0.6558	0.0218	0.0189
LCS	-0.4999	16.9674	0.0218	0.5060
OLS	-0.4999	16.9674	0.0218	0.5060
PLRT	9.7718	11.6344	0.8011	0.7947
RTM	-0.4999	16.9674	0.0218	0.5060

The comparative performance of change estimation models demonstrates that bias and error must be interpreted relative to the scale of measurement. On the ordinal $[1, 7]$ scale, the piecewise linear regression corridor model (PLRC) shows a notable advantage in bias reduction. Its mean bias of $+0.1847$ corresponds to approximately 3% of the scale range, compared to -0.4999 for the beta regression model—nearly 8% of the total ordinal range. Although beta regression achieves the lowest RMSE (0.6550), PLRC remains competitive (0.6874). These results suggest that PLRC is well-suited for discrete, bounded scales such as Likert-type items.

However, this advantage does not extend to continuous-scale settings. When applied to responses transformed onto a $[0, 100]$ visual analog scale (VAS), PLRC exhibits substantial overestimation, with a mean bias of $+9.77$—roughly 10% of the total scale. In contrast, beta regression maintains a nearly constant bias of -0.4999 (less than 1% of the range), along with lower RMSE and variability. This contrast highlights the scale sensitivity of PLRC: while its thresholded, piecewise structure is advantageous in ordinal contexts, it introduces bias and instability under smoothly varying, quasi-continuous conditions.

Other benchmark models, including OLS, LCS, and RTM, perform similarly across simulations, largely due to their shared reliance on continuous, unbounded functional forms. Their performance aligns more closely with continuous response formats, such VAS, but they lack the capacity to capture asymmetric slope regimes. In this respect, the PLRC model fills a methodological niche by explicitly modeling these asymmetric regimes and approximating nonlinear response dynamics

through threshold-defined linear segments.

While PLRC may be less efficient in continuous contexts, its structural alignment with ordinal measurement makes it a practical tool for modeling discrete attitude shifts. To further illustrate its utility, we extend PLRC to incorporate individual-level covariates and apply it to empirical data, returning to Vlasceanu et al. (2021) to examine the role of behavior-level predictors in modulating response change.

2.8 Applying PLRC

PLRC leverages structural constraints to explain behavioral variation, shifting the analytic focus from inferring latent dispositions to examining how measurement structure shapes responses. Again, this represents a methodological pivot: from interpreting responses as reflections of internal states to viewing them as co-determined by individual traits and structural affordances.

To isolate person-level responsiveness, we introduce a leave-one-out *relative responsiveness index*:

$$\tilde{\rho}_{ji}^{\text{loo}} = \frac{1}{I-1} \sum_{k \neq i} \frac{|\Delta T_{jk}|}{\max(|s - T_{jk}^{1}|, \varepsilon)} \tag{2}$$

where ΔT_{jk} is the observed change on the item k for participant j, T_{jk}^{1} is that person's baseline, s is the evidence-aligned target set to 0 for refutative evidence or 100 for supportive, and ε is a small constant to prevent division by zero in cases where $T_{jk}^{1} = s$, and I is the total number of items (36 propositions). This is change per unit of potential movement—how much someone moves relative to the room they had to move. This index captures the normalized responsiveness of individual j across all *other items*, excluding the current item i. Larger values of $\tilde{\rho}$ indicate the participant changes a lot on other items — they're generally responsive. If low, they're more inert — typically unresponsive.

The linear mixed-effects model estimating change $\Delta T_{ji} = T_{ji}^{2} - T_{ji}^{1}$ is specified as:

$$\Delta T_{ji} = \beta_1 (s - T_{ji}^{1}) + \beta_2 \tilde{\rho}_{ji}^{\text{loo}} + \beta_3 d_{L,ji} + \beta_4 d_{U,ji} + u_j + v_i + \varepsilon_{ji} \tag{5}$$

with the components defined as:

$$d_L = \max\big(L - T^1, 0\big), \text{(active when } T^1 < L\text{)}$$

$$d_U = \max\big(T^1 - U, 0\big), \text{(active when } T^1 > U\text{)}$$

$$u_j,\ v_i = \text{random effects for participant and item, respectively,}$$

$$\varepsilon_{ij} = \text{residual error.}$$

This model partitions the space into three fixed-effect behavioral regimes for:

Region	**Change equation**
$L \le T^1 \le U$	$\Delta T = \beta_1(s - T^1) + \beta_2\tilde{\rho}^{\text{loo}}$
$T^1 < L$	$\Delta T = \beta_1(s - T^1) + \beta_2\tilde{\rho}^{\text{loo}} + \beta_3 d_L$
$T^1 > U$	$\Delta T = \beta_1(s - T^1) + \beta_2\tilde{\rho}^{\text{loo}} + \beta_4 d_U$

The model does not choose between threshold induced deviation and evidence; rather, both can operate simultaneously. If both deviation and evidence push in the same direction, their effects compound; if they oppose, they compete. The responsiveness term modulates how sensitive individuals are to these forces. Critically, when $T^1_{ij} \in [L, U]$, both deviation components d_L and d_U are zero, and change is modeled as a function of evidence discrepancy and normalized change relative to other items. Outside the corridor, however, the deviation terms operate as structural corrections that adjust for systematic directional change tied to baseline extremity. This prevents the model from over-attributing movement near the ends of the scale to the effects of evidence alone.

More broadly, the deviation mechanism serves multiple functions: it encodes theoretical expectations about bounded response behavior; it imposes a regularizing effect by attenuating noise-driven deviations near the scale extremes; and it enhances the interpretability of evidence effects by disentangling evidence–baseline discrepancy from scale–baseline proximity.

We test this model using data from Vlasceanu et al. (2021). The results are presented in Table

13. The fixed effects estimates characterize how evidence discrepancy, boundary-proximity terms, and individual responsiveness covary with observed change scores (ΔT).

Table 13: Fixed Effects Estimates: L = 25, U = 75

Term	Estimate	Std. Error	p-value	95% CI Lower	95% CI Upper
$(s - T^1)$	1.7072	0.0079	$< 2.2 \times 10^{-16}$	1.6916	1.7228
$\tilde{\rho}^{loo}$	−0.0735	0.0477	0.1234	−0.1671	0.0200
d_L	0.2383	0.0260	4.77×10^{-20}	0.1873	0.2892
d_U	−0.0884	0.0212	3.08×10^{-5}	−0.1300	−0.0468

The coefficient on the evidence discrepancy term $(s - T^1)$ is the largest in magnitude and significant ($\hat{\beta}_1 = 1.707$, $p < .001$). This suggests that changes in ΔT are systematically larger when the discrepancy between the evidence-consistent target s and the initial score T^1 is greater. This is consistent with the model's structure, which scales the potential for updating according to the available distance between baseline response and external evidence.[11]

The responsiveness term $\tilde{\rho}^{\text{loo}}$—reflecting the participant's average responsiveness on other items—is negative but not statistically significant ($\hat{\beta}_2 = -0.0735$, $p = 0.123$). While the direction is aligned with theoretical expectations (participants who typically exhibit large shifts elsewhere may require less model-based correction on a given item), this effect is not reliably different from zero in this specification. Its interpretation as a trait-like moderator of sensitivity remains viable, though its statistical contribution here appears limited.

The lower-bound discrepancy term $d_L = L - T^1$ is significant and positively signed ($\hat{\beta}_3 = 0.238$, $p < .001$). Since this term is activated when T^1 is near the lower end of the response scale ($T^1 < L$), the positive coefficient corresponds to larger shifts upward from low-scoring initial positions.

[11]The random effects estimates reveal notable heterogeneity at both the participant and item levels. The standard deviation for `participant_group` is 6.17, indicating modest variation in baseline change behavior across individuals (and partisan lines). In contrast, the standard deviation for `item_group` is considerably larger at 15.93, suggesting that items differ more substantially in the extent of change they induce. The residual standard deviation of 25.75 underscores the presence of significant unexplained variability—common in survey data.

The upper-bound term $d_U = T^1 - U$ is negative and significant ($\hat{\beta}_4 = -0.0884$, $p < .001$). This term activates when T^1 is near the upper bound ($T^1 > U$), which typically corresponds to cases shifts are expected to be downward. Since d_U is itself positive in this region, the negative coefficient results in a net negative product: ($\beta \cdot d_U < 0$), thus reinforcing downward shifts when initial scores are high.

Collinearity among fixed effects is generally low, with most pairwise correlations falling below $|r| = 0.26$. Two notable exceptions involve structurally related terms—specifically, between $s - T^1$ and d_L ($r = 0.715$), and between $s - T^1$ and d_U ($r = -0.759$).

Table 14: Correlation Matrix of Fixed Effects

	$s - T^1$	$\tilde{\rho}^{loo}$	d_L
$\tilde{\rho}^{loo}$	-0.011		
d_L	0.715	-0.037	
d_U	-0.759	-0.035	-0.426

First, the observed positive correlation between the evidence discrepancy term ($s - T^1$) and the lower-bound deviation term $d_L = L - T^1$ ($r = 0.715$) reflects their systematic co-movement in the lower region of the response scale ($T^1 < L$), despite having opposite signs. When the evidence target is fixed at $s = 0$, the discrepancy term ($s - T^1$) is strictly negative, while d_L is strictly positive. As T^1 decreases toward the bottom of the scale, d_L increases, and ($s - T^1$) becomes less negative—increasing numerically toward zero. Conversely, as T^1 increases toward L, both d_L and ($s - T^1$) decrease: d_L shrinks, and ($s - T^1$) becomes more negative. Although these terms differ in sign, they vary in the same numerical direction as T^1 changes, resulting in a strong positive correlation. This relationship arises from their shared dependence on T^1.

Conversely, the upper-bound affordance term $d_U = T^1 - U$ exhibits a strong negative correlation with the evidence discrepancy term ($s - T^1$) ($r = -0.759$), driven by their opposing directional dependence on T^1 in the upper range of the scale ($T^1 > U$). When $s = 0$ and U is fixed (e.g., 75),

increasing T^1 simultaneously increases d_U and decreases $(s - T^1)$, as initial responses move farther above the upper threshold, d_U grows positively, while the discrepancy from the evidence target becomes more negative. A similar pattern holds for $s = 100$: as T^1 increases, both d_U and $(s - T^1)$ move in opposite numerical directions. As T^1 decreases toward the upper threshold U from above $(T^1 > U)$, the affordance term $d_U = T^1 - U$ shrinks toward zero, the evidence discrepancy term $s - T^1$ simultaneously increases (becomes less negative) assuming a fixed target s. Thus, even as T^1 approaches U, the two terms continue to vary in opposing directions: d_U decreases, while $s - T^1$ increases. This reinforces the negative correlation between these structurally defined predictors across the relevant domain.

This antagonistic co-movement produces the negative correlation observed in the data and reflects the inherent structure of deviation-discrepancy interactions under bounded conditions. These empirical correlation patterns reveal a deeper invariant structural feature of bounded updating processes in two-wave survey designs. Specifically, the observed relationships

$$\text{corr}(d_U,\ s - T^1) < 0 \quad \text{and} \quad \text{corr}(d_L,\ s - T^1) > 0$$

hold invariantly across polar evidence conditions (e.g., $s = 0$ or $s = 100$). *This reflects an asymmetry built into the geometry of bounded opinion scales: the discrepancy between a respondent's initial attitude T^1 and an external target s are inherently entangled irrespective of the direction of evidence.*

The sign and strength of these correlations are therefore not artifacts of the specific evidence introduced but rather consequences of how deviation terms and discrepancy values co-vary across the bounded response space. These findings provide a structural explanation for the moderating role of boundary terms in two-wave response models. The associations between d_L, d_U, and $s - T^1$ arise from intrinsic properties of the bounded response scale, rather than from the substantive content or direction of the intervention. As such, these terms function as stable statistical moderators due to their geometric relationship to response potential, not because of any specific informational effect.

Consequently, informational interventions may appear to have asymmetric effects even when the

evidence is symmetric in polarity, due to differences in the structural interaction between evidence discrepancy and baseline position.

Most prior survey or belief change models do not account for the interaction between boundary proximity and perceived discrepancy in an explicit structural manner. The concept that the scale geometry itself imposes structural asymmetries—which then moderate responsiveness to interventions—is not typically formalized in mainstream longitudinal or attitude change models. Our framing elevates boundary proximity to a systematic, modelable influence that persists under transformation of the evidence.

Summing up the PLRC to this point, the model allows distinctions to be drawn between learning-like processes (i.e., evidence discrepancy) and structural influences introduced by bounded response scales.

It explicitly distinguishes between:

- *Evidence-Discrepancy*, operationalized through the discrepancy between the baseline score and the evidence-consistent target;
- *Responsiveness*, captured via a leave-one-out index that quantifies individual sensitivity to change across other items;
- *Threshold Deviation*, represented as a structural component that modifies predicted change outside a central interval, depending on proximity to threshold points and bounds.

By allowing these mechanisms to operate concurrently, the model explicitly incorporates structural correction via the piecewise deviation terms d_L and d_U, which are activated only when scores fall outside a defined neutral corridor. When an initial score lies near the floor or ceiling of the scale, these terms exert directional influence, and depending on the coefficient β can capture phenomena that may resemble ceiling/floor effects or overshoot. The interaction between the deviation terms and the evidence discrepancy term enables the model to express asymmetric updating: when structural correction and evidence direction are aligned, the resulting change is

amplified; when they are opposed, the net effect is attenuated. This formulation allows the model to account for observed heterogeneity in response change without attributing it solely to individual differences in response or error.

Further, by incorporating a participant-level responsiveness index $\tilde{\rho}$, the model captures individual differences in the propensity to revise beliefs. This allows the model to distinguish between similar endline responses that result from different factors, thereby reducing overgeneralization in attributing predicted change to specific features.

Altogether, the model's structure supports the examination of: (1) how response change aligns (or misaligns) with evidence-based discrepancies; (2) the potential influence of structural features introduced by bounded response scales; and (3) systematic variation in individual responsiveness. This integrated framework offers a theoretically motivated and empirically workable approach to a long-standing challenge in behavioral research—characterizing response dynamics within bounded survey formats.

To assess multicollinearity, we report Variance Inflation Factors (VIFs) from the fixed effects model: $(s - R_1) = 4.76, \quad \tilde{\rho}^{loo} = 1.00, \quad d_L = 2.36, \quad d_U = 2.84$. These values fall well below conventional thresholds, indicating no concerning multicollinearity. Accordingly, we proceed with the interaction model (Table 15), where structural asymmetries remain evident (Table 16).

Table 15: Linear Mixed-Effects Model Estimates
Model: $\Delta T_{ji} \sim 0 + (s - T^1)\ \tilde{\rho}^{loo} + (s - T^1)\ d_L + (s - T^1)\ d_U + \tilde{\rho}^{loo}\ d_L + \tilde{\rho}^{loo}\ d_U + (1 \mid \text{participant\_group}) + (1 \mid \text{item\_group})$ (6)

Fixed Effect	**Estimate**	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
$(s - T^1)$	1.690	0.0154	63,820	109.94	$< 2e{-}16^{***}$
$\tilde{\rho}^{loo}$	–0.111	0.0518	2,300	–2.14	0.0326^{*}
d_L	–0.396	0.1758	63,920	–2.25	0.0244^{*}
d_U	0.248	0.1323	63,720	1.88	$0.0605^{\cdot}$
$(s - T^1){:}\tilde{\rho}^{loo}$	–0.00037	0.00249	63,690	–0.15	0.8827
$(s - T^1){:}d_L$	0.01398	0.00341	63,930	4.10	$4.23e{-}05^{***}$
$(s - T^1){:}d_U$	0.00624	0.00247	63,430	2.53	0.0115^{*}
$\tilde{\rho}^{loo}{:}d_L$	–0.0141	0.00773	63,230	–1.83	$0.0676^{\cdot}$
$\tilde{\rho}^{loo}{:}d_U$	–0.00021	0.00657	62,800	–0.03	0.9743

Model Fit: REML criterion at convergence: 599522.6

Significance codes — $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 16: Correlation Matrix of Fixed Effects (Interaction Model)

	$s - T^1$	$\tilde{\rho}^{loo}$	d_L	d_U	$\tilde{\rho}^{loo}{:}d_L$	$s - T^1{:}d_L$	$s - T^1{:}d_U$	$\tilde{\rho}^{loo}{:}d_L$
$\tilde{\rho}^{loo}$	-0.099							
d_L	0.406	-0.087						
d_U	-0.455	-0.055	-0.147					
$s - T^1{:}\tilde{\rho}^{loo}$	-0.809	0.117	-0.198	0.229				
$s - T^1{:}d_L$	-0.222	0.028	-0.965	0.080	0.036			
$s - T^1{:}d_U$	-0.240	-0.029	-0.078	0.956	0.035	0.043		
$\tilde{\rho}^{loo}{:}d_L$	-0.604	0.258	-0.255	0.134	0.745	0.043	0.014	
$\tilde{\rho}^{loo}{:}d_U$	0.647	0.103	0.124	-0.292	-0.797	-0.017	-0.054	-.496

The correlation matrix reveals several structurally meaningful relationships among the fixed effects, particularly those involving the deviation terms d_L and d_U, and their interactions with the

evidence discrepancy term, approximated by $s - T^1$. As shown previously, d_L exhibits a moderate positive correlation with $s - T^1$ ($r = 0.406$), while d_U shows a moderate negative correlation with $s - T^1$ ($r = -0.455$), consistent with their respective activation near the lower and upper bounds of the scale. These opposing signs again reflect the directional geometry of bounded response behavior: as initial scores move away from the evidence target, proximity to the scale boundaries introduces deviation effects that either amplify or attenuate predicted change. Importantly, the interaction terms $(s - T^1) : d_L$ and $(s - T^1) : d_U$ display strong correlations with their corresponding deviation terms—$r = -0.965$ for $(s - T^1) : d_L$, and $r = 0.956$ for $(s - T^1) : d_U$—highlighting likely multicollinearity. While these interaction terms are only weakly correlated with $s - T^1$ itself ($r = -0.222$ and $r = -0.240$, respectively), this reflects a nonlinear clipping effect that weakens the correlation. Nevertheless, their strong alignment with the boundary terms confirms that they capture local, context-dependent amplification or suppression of evidence-based updating. Collectively, these correlations preserve the model's capacity to distinguish between updating driven by global evidence discrepancy and that which emerges from structural effects—thus maintaining interpretability even as model complexity increases.

Empirically, the results support the conclusion that the geometry of the response scale plays a meaningful—potentially substantial—role in shaping response change. Specifically, both the distance from the target value and the proximity of initial scores to the upper and lower threshold points influence adjustment behavior. These geometric constraints do not operate independently; rather, they modulate the expression of evidence discrepancy, indicating that response dynamics on bounded scales arise from the joint influence of informational signals and structural positioning.

2.9 A Different Perspective

Traditional psychometrics treats items (i.e., individual survey questions or prompts) as imperfect indicators of an underlying latent trait, assuming the response scale operates as a neutral measurement channel. In contrast, the PLRC model inverts this perspective, emphasizing the structural influence of the response format itself on observed responses. It treats the scale itself as an eco-

logical structure—an environment with affordances and constraints that shape, enable, or inhibit movement. Observed belief change is modeled as a function of three primary components here: (i) an evidence-based component proportional to the discrepancy between the evidence target and the initial response ($s - T^1$); (ii) structural deviation terms d_L and d_U, which are activated only when the initial response lies outside the predefined stability corridor; (iii) an individualized responsiveness metric $\hat{\rho}^{loo}$. By accounting for structural effects tied to boundary proximity, PLRC helps disentangle evidence sensitivity estimates from evidence-independent change induced by the bounds of the response scale.

Methodologically, the PLRC model—to our knowledge—offers the first unified and empirically testable framework that simultaneously integrates evidence discrepancy, piecewise structurally induced deviation, and person-level responsiveness within a single model. Rather than treating observed responses solely as reflections of latent traits, PLRC reconceptualizes belief change as the joint product of interactions among items, individual responsiveness, and the structural affordances imposed by the measurement instrument itself.

The interpretability of PLRC depends on several key assumptions. First, the functional form must be valid: deviation is assumed to operate piecewise-linearly outside a fixed neutral zone $[L, U]$, while the effect of evidence discrepancy is assumed to be constant across the scale. Second, parameters must be identifiable. Although collinearity between $s - T^1$ and the deviation terms is structurally induced by the bounded scale, it compromises precision (via inflated standard errors) rather than the validity of the parameter decomposition. Diagnostic tools such as variance inflation factors or posterior correlations can assess this issue. Mitigation strategies include residualizing deviation terms on the evidence discrepancy or estimating the evidence effect solely within the stability corridor and deviation effects outside it.

Third, evidence stimuli must be equivalent across groups; otherwise, group differences in $s - T^1$ may conflate variation in stimulus semantics with structural features of the response process. Fourth, sufficient data density near the scale's boundaries is necessary to estimate d_L and d_U reliably. Each

of these assumptions is empirically testable.

Specifications, Assumptions, Limitations, and Reasoning

The mixed-effects model is specified to reflect the hierarchical structure of behavioral data, following best practices outlined by Raudenbush and Bryk (2002). First, the model accounts for the crossed data structure by including random intercepts for both participants and items, recognizing that responses are nested within both units. Second, it addresses non-independence within participants and items by explicitly modeling these dependencies through random effects, ensuring valid estimation of standard errors. Third, it accommodates heterogeneous baselines across individuals and items, allowing for variability in average response tendencies through the inclusion of random intercepts. Fourth, the model remains robust to unbalanced data—a common feature in behavioral research—by leveraging the flexibility of mixed-effects estimation, which does not require complete data across all units. Fifth, to address the bounded nature of the outcome variable (responses on a 0–100 scale), the model incorporates structural deviation terms (d_L, d_U) that capture boundary-related contributions to response change. These terms reflect the influence of proximity to the lower and upper bounds of the scale, allowing the model to adjust for geometric constraints inherent to bounded measurement. Finally, the inclusion of theoretically motivated interaction terms—for example, between prediction error and deviation or between evidence discrepancy and responsiveness—enables the model to capture context-sensitive, non-linear response patterns. This enhances both the interpretability and theoretical alignment of the model, supporting a more nuanced understanding of adaptive response change.

Survey responses are modeled as numeric values on a fixed, bounded continuum (e.g., [0, 100]), where each unit represents a meaningful, equal interval. As discussed, observed changes are interpreted in part as a function of available "headroom"—the remaining space on the scale—embedding a spatial or affordance-based logic into the response process. This variation may be moderated by evidence type and general responsiveness to other items.

The model adopts a piecewise linear specification: boundary-related terms activate when

responses fall outside a central corridor, with linear correction applied within each regime. However, real-world boundary effects may be nonlinear—e.g., diminishing influence with greater distance from the bound—leading to potential attenuation or amplification of structural effects across the scale.

Random intercepts for both participants and items account for within-cluster dependence. Future extensions could incorporate random slopes or richer residual structures to capture within-person heterogeneity more fully.

This approach departs from standard latent-change models, which typically attribute updates to evidence under assumptions of homogeneous error. Instead, it models change as jointly driven by behavioral variation and structural properties of the response scale. *This may help explain why roughly 40% of observed updates appear incongruent under purely evidence-based models, underscoring the extent to which response scale geometry can shape observed patterns of change. While such patterns are often treated as anomalies within conventional analytic frameworks, we argue that they reflect systematic, modelable dynamics.* The following section advances this modeling approach using data from Study 2.

3. Cross-Method (Likert) Comparison

3.1 Data and Source (Siegrist & Bearth, 2021)

We utilize a two-wave panel study administered in Switzerland in 2020, comprising 1,223 adults and assessing their perceptions of COVID-19 risk. Risk perception was measured across five-items using a 1 (no fear) to 7 (high fear) discrete integer Likert scale, capturing concerns such as fatality and hospital strain. This yields 6165 observations of change. The two waves—spaced by the ongoing progression of the SARS-CoV-2 pandemic—were administered from March 27 to April 5 and April 17 to April 26, respectively, with the same participants. *In this study, no evidence or intervention was introduced; only the passage of time separates the two measurement waves.* While the original study examined perceived risk and the acceptance of policy interventions, our analysis

isolates and draws exclusively on the perceived risk component.

Figure 9 presents the average response change across seven bins, each representing a one-point increment along the response scale, for a single proposition from the administered set. Tile colors encode the mean change from Wave 1 to Wave 2 (ΔT): green indicates a positive shift (increase), red indicates a negative shift (decrease), and white denotes a negligible change. Superimposed arrows represent the estimated local slope $\hat{\rho}$; their direction reflects the sign of change, and their length denotes relative magnitude. Importantly, no evidence or intervention was introduced between waves for this dataset—the changes observed reflect only temporal dynamics across measurement occasions.

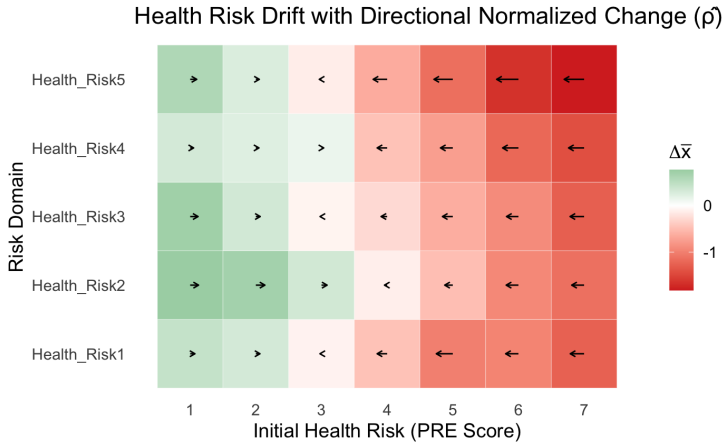


Figure 9: Results from Source: Siegrist & Bearth (2021)

The observed pattern superficially resembles RTM. However, the magnitude of observed change is asymmetrical: respondents starting at the high end of the belief scale (e.g., bin 7) show substantially larger average shifts (–1.8) in magnitude compared to the modest shifts observed among low-end respondents (e.g., +0.4 for bin 1), despite both positions being equidistant from the scale's midpoint. In addition, normalized deviation values ($\hat{\rho}$) range from –0.20 to –0.33 in upper bins (5 to 7), but only +0.05 to +0.13 in lower bins (1 to 3), indicating an asymmetry in both average adjustment direction and extent.

We evaluate whether observed response change is best explained under a RTM framework by statistically comparing the empirical data to a bounded, noise-based RTM null model using the

same number of bins.

3.2 RTM?

We construct a bounded RTM null model by simulating response change as normally distributed noise around each participant's observed pre-test score, constrained to the original 1–7 scale. Specifically, simulated follow-up scores were generated as:

$$T^2_{\text{sim}} = T^1 + \varepsilon, \quad \varepsilon \sim N(0, \sigma)$$

This simulation was repeated 1,000 times, producing a distribution of expected change under RTM alone. For each bin of initial response and each health risk item, we computed the mean simulated change and corresponding confidence intervals. These results serve as a benchmark for detecting systematic deviations from RTM in the empirical data.

Next, we compared observed change ($\Delta T = T^2 - T^1$) to the simulated null using Cohen's d, defined as:

$$d = \frac{\text{Observed} - \text{Expected}}{\text{Standard Deviation of Null}}$$

This represents a one-sample effect size comparing the observed mean change to the expected mean under the null distribution.[12]

This comparison reveals that while the RTM model predicts a shallow, symmetric deviation pattern centered around the scale midpoint, the empirical data exhibit a steep and asymmetric pattern of change—indicating once again, that RTM alone cannot account for the observed updating behavior.

Figure 10 displays the expected RTM trajectories for each risk perception item, with grey bands

[12] If comparing two independent empirical samples, it would be appropriate to use the pooled variance: $s_p = \sqrt{\frac{s_1^2 + s_2^2}{2}}$. However, in this analysis, the observed mean is fixed, and only the null distribution includes variability. Therefore, the denominator uses the standard deviation of the simulated null distribution alone.

representing 95% confidence intervals derived from 1,000 simulated draws under the RTM null model. The blue lines show the observed mean changes, estimated via finite differences across bins of initial response scores.

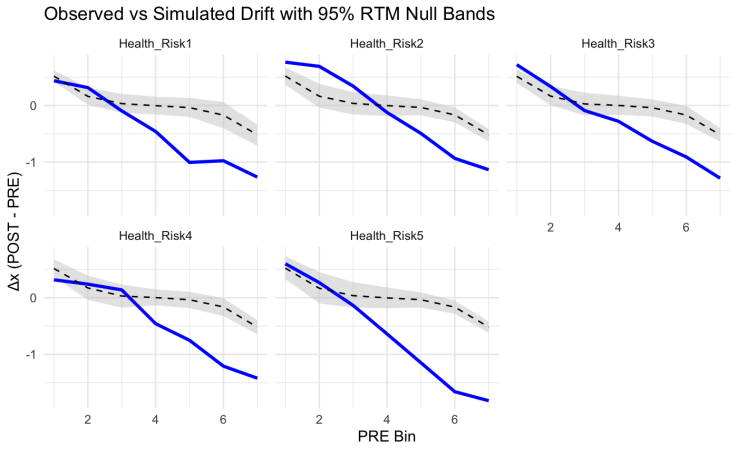


Figure 10: Data from Siegrist & Bearth (2021)

To further examine these dynamics, we conducted a slope analysis centered on the observed inflection point (bin 4), applying LOESS smoothing to estimate local trends. We compared the slopes of change immediately before and after this threshold. In all domains, the post-inflection slope became sharply negative, diverging from the shallow, monotonic gradient predicted under classical RTM. This pattern provides additional evidence for a bounded asymmetry effect. Finally, we implemented a bootstrap-based inference procedure to compute empirical p-values for each bin, estimating the proportion of simulated RTM trajectories that produced changes more extreme than those observed. The results confirmed that changes in the upper bins significantly deviated in both magnitude and direction from RTM expectations.

Table 17: Change in Slope around Midpoint 4 and Significance by Health Risk Domain

Item	Slope (Before)	Slope (After)	p-val (Before)	p-val (After)	Slope Diff
Health_Risk1	-0.266	-0.239	0.194	0.086	0.027
Health_Risk2	-0.212	-0.347	0.229	0.010	-0.136
Health_Risk3	-0.405	-0.329	0.018	0.002	0.076
Health_Risk4	-0.089	-0.335	0.051	0.009	-0.245
Health_Risk5	-0.368	-0.405	0.035	0.022	-0.037

Table 17 summarizes changes in response slopes around the midpoint bin (T^1 =4) across five health risk items. In each domain, we observe systematic variation in the direction and magnitude of response change before and after the inflection point. Specifically, the slope becomes more negative post-midpoint in four of five items, with statistically significant shifts ($p < .05$) observed in the post-inflection region for all but Health_Risk1. This indicates that participants are more likely to exhibit downward adjustments when starting above the midpoint, compared to those starting below it–as open space theory would lead us to expect. The consistent shift in slope direction and significance—particularly the steeper decline in upper bins—suggests that response change is not symmetric with respect to the scale's center. Instead, the descriptive results suggest three interpretations: (i) observed change magnitudes are asymmetric, (ii) responses become more negative as the initial value increases, and (iii) these deviations significantly diverge from those predicted by a noise-based RTM null model.

To evaluate whether the magnitude of response change varies systematically with respondents' initial position on the 7-point scale, we conducted a non-parametric permutation test.

3.3 Permutation Test

To assess whether observed response change is systematically related to baseline response level, we performed a permutation test using bin-specific mean change as the target statistic.

Let the observed change for participant j on item i be:

$$\Delta T_{ji} = T_{ji}^2 - T_{ji}^1,$$

where $T_{ji}^1 \in \{1,\ldots,7\}$ is the baseline response, and T_{ji}^2 is the follow-up response.

We computed the average change within each initial response bin $r \in \{1,\ldots,7\}$:

$$\bar{\Delta}_r = \mathbb{E}[\Delta_{ji} \mid T_{ji}^1 = r],$$

where $\bar{\Delta}_r$ is the empirical mean change for all respondents whose baseline score was r.

To summarize heterogeneity across bins, we calculated the variance of the bin means:

$$\mathrm{Var}(\bar{\Delta}) = \frac{1}{6}\sum_{r=1}^{7}\left(\bar{\Delta}_r - \overline{\bar{\Delta}}\right)^2, \quad \text{where } \overline{\bar{\Delta}} = \frac{1}{7}\sum_{r=1}^{7}\bar{\Delta}_r.$$

To generate a null distribution under the hypothesis

$$H_0 : \bar{\Delta}_1 = \bar{\Delta}_2 = \cdots = \bar{\Delta}_7,$$

we permuted baseline responses T_{ji}^1 across participants j for each item i, while keeping ΔT_{ji} fixed. This preserves the marginal distribution of change scores while removing any dependence on initial response bin. For each permutation, we recalculated $\mathrm{Var}(\bar{\Delta})$, yielding a distribution of surrogate variance values under the null. We repeated this process $N = 5{,}000$ times per item, producing surrogate values $\mathrm{Var}^{*(k)}$. The empirical p-value is given by:

$$\hat{p} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{1}\left\{\mathrm{Var}^{*(k)} \geq \mathrm{Var}(\bar{\Delta})\right\},$$

where $\mathbf{1}\{\cdot\}$ is the indicator function.

Table 18: Permutation Test Results by Risk Item

Risk Item	Observed Var($\bar{\Delta}$)	Empirical p-value
Health_Risk1	0.458	< 0.0002
Health_Risk2	0.581	< 0.0002
Health_Risk3	0.490	< 0.0002
Health_Risk4	0.503	< 0.0002
Health_Risk5	0.884	< 0.0002

For all five health-risk items, the observed bin-level variance in response change exceeded the maximum variance observed in any of the 5,000 permutations, yielding empirical p-values < 0.0002. This provides strong evidence that the magnitude of response change varies systematically with the respondent's initial position on the scale.

Substantively, these findings again suggest that initial placement on a bounded response scale plays a consequential role in shaping both the direction and magnitude of response change. The effect is most pronounced for Health_Risk5 (Var $= 0.884$), suggesting greater baseline-driven heterogeneity in that domain. This supports the broader claim that bounded measurement instruments can induce structural asymmetries in longitudinal data—even in the absence of differential evidence content—by conditioning the range and direction of possible updates.

3.4 PLRC as a Forecasting Method

We now use the PLRC model as a generative framework here to simulate expected response shifts and evaluate fit against the observed change distribution. To simulate response behavior, each respondent–item pair is assigned a latent baseline $Y_{ji} \sim U\{1,2,\ldots,7\}$, representing the respondent's "true" score on the bounded 7-point scale. This value is treated as the initial position from which directional change is computed. The PLRC corridor is defined here as $[L,U] = [3,5]$. This structure functions as a directional analog to regression to the mean, whereby simulated respondents with

baselines outside the corridor are systematically drawn back toward its interior.

Expected change for respondent j on item i is defined by a piecewise-linear function:

$$\Delta T_{ji} = \alpha + \begin{cases} \beta_L \cdot (L - Y_{ji}), & \text{if } Y_{ji} < L \\ 0, & \text{if } L \leq Y_{ji} \leq U \\ \beta_U \cdot (Y_{ji} - U), & \text{if } Y_{ji} > U \end{cases}$$

where α is a global intercept, and β_L, β_U are slope parameters governing the strength of the correction below and above the corridor, respectively.

To map these continuous predictions to discrete change outcomes (ranging from -6 to $+6$), a Gaussian smoothing procedure is applied:

$$X_{ji} \sim N(\mu_{ji}, \sigma^2), \quad \Delta_{ji} = \text{round}(X_{ji}),$$

where $\mu_{ji} = \Delta T_{ji}$. Rounding yields integer-valued change scores, and the smoothed distribution over the 13 discrete outcomes is used to evaluate model fit.

Model parameters are estimated using penalized maximum likelihood by minimizing the multinomial deviance between predicted and empirical change frequencies. This approach directly assesses how well the model, informed by structural scale geometry, reproduces observed patterns of change.

Parameter	β_L	β_U	α	σ
Estimate	0.500	0.800	-0.526	1.487

RMSE = 0.0314 ($\approx$ 3.1% per bin)

Table 19: Model fit summary: Estimated parameters and fit statistics for the PLRC model.

The estimated parameters of the PLRC model (Table 19) reveal asymmetric corrective dynamics across the response scale. The lower-bound slope, $\beta_L = 0.500$, reflects a moderate tendency for upward adjustment among respondents whose initial scores fall below the lower threshold ($S_{ji} < L$). In contrast, the steeper upper-bound slope, $\beta_U = 0.800$, indicates a stronger corrective effect for scores exceeding the upper threshold ($S_{ji} > U$). This asymmetry partly arises from the configuration of the corridor itself: with $U = 5$, the only possible upper-bound exceedances occur at response levels 6 or 7. As a result, the discrepancy term $Y_{ji} - U$ takes on values of 1 or 2, requiring a relatively large positive coefficient to match the observed magnitude of change given the estimated global intercept, $\alpha = -0.526$. This intercept introduces a uniform downward shift across the full range of baseline responses. Accordingly, even among high scorers with a positive upper-slope effect, the net predicted change may remain negative. The noise parameter, $\sigma = 1.487$, captures moderate residual variation around the predicted values, consistent with the model's Gaussian error structure. Model fit is evaluated by comparing predicted and observed change frequencies across 13 discrete bins (Table 2). The root mean square error (RMSE) of 0.0314 ($\approx 3.1\%$ per bin) indicates a strong overall fit. As shown in Figure 11, the model closely tracks the empirical distribution across most bins.

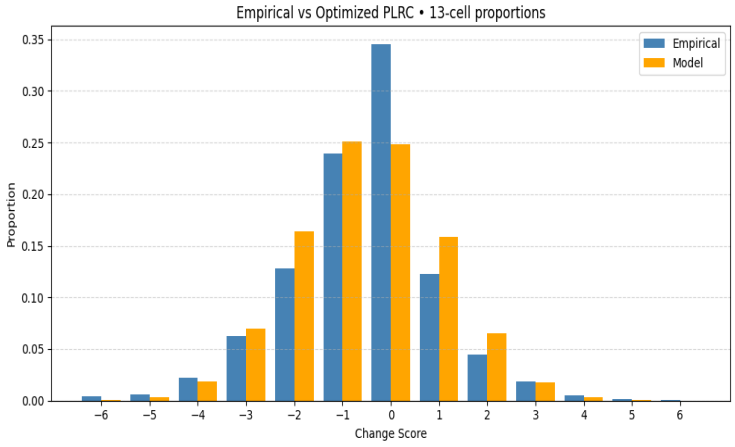


Figure 11: Empirical data (blue)–Health, and PLRC (orange).

Change	Empirical	Model	Difference
-6	0.0041	0.0004	0.0037
-5	0.0056	0.0033	0.0022
-4	0.0224	0.0190	0.0034
-3	0.0626	0.0694	-0.0068
-2	0.1279	0.1641	-0.0362
-1	0.2389	0.2508	-0.0119
0	0.3455	0.2480	0.0976
1	0.1228	0.1586	-0.0358
2	0.0445	0.0656	-0.0211
3	0.0185	0.0175	0.0010
4	0.0051	0.0030	0.0020
5	0.0015	0.0003	0.0011
6	0.0007	0.0000	0.0006

Table 20: Comparison of empirical and model-predicted change distributions across 13 bins.

Notably, the empirical data exhibit a pronounced central spike at zero change (34.5%), which the model underestimates (24.8%). This suggests that many respondents exhibit complete score stability between waves—potentially due to true attitudinal consistency, response inertia, or anchoring effects. The PLRC model, by contrast, assumes more consistent deviation modulated by latent baselines and scale geometry, producing a ”smoother” distribution centered at zero. This mismatch implies that the model lacks mechanisms to capture resistance to change or consistency in response.

The model tends to slightly overpredict modest positive changes (e.g., $+1$, $+2$), reflecting the relatively strong correction slope assigned to respondents with high initial scores. This pattern is consistent with the idea that structural asymmetries in the response scale manifest differently across its range. However, the model's omission of a latent, state-dependent inertia component may limit

its capacity to fully capture respondent behavior near the midpoint of the scale, where changes are typically smaller and less directional.

These findings underscore the now familiar core premise of the PLRC framework: that the geometry of the response scale—including its boundaries and central corridor—can systematically structure observed change. Specifically, it reproduces three key features of the empirical distribution: (1) directional asymmetry, with downward correction among high scorers but little upward movement among low scorers; (2) a central peak consistent with inertia or resistance near the corridor; and (3) a modest global drift, reflecting aggregate bias. Together, these patterns suggest that operationalizing just structural features of the scale can account for a substantial portion of the observed response dynamics, independent of any evidence.

Source	N	**Mean Δ**	**Median Δ**	**SD Δ**	**Skewness**	**Kurtosis**
Empirical	6,115	−0.526	0	1.514	−0.167	4.224
Simulated	6,115	−0.203	0	1.087	−0.264	2.957
RMSE of change–score proportions = 0.032 (3.2% per bin)						

Table 21: Summary statistics for empirical and PLRC-simulated change-score distributions, with aggregate fit diagnostics.

Using a three-parameter mean function with homoscedastic Gaussian noise, the PLRC model achieves strong agreement with empirical data: a root mean squared error (RMSE) of 3.2% across 13 discrete outcome bins. Parameters were estimated via penalized maximum likelihood, minimizing the deviance between observed and predicted change-score frequencies. Despite its simplicity, the PLRC captures structural properties of the empirical distribution. Specifically, it:

- Accurately reproduces the sign reversal in deviation at the corridor cut-points L=3 and U=5;
- Generates asymmetric correction pressure—downward for high scorers, neutral for low scorers;

- Accounts for roughly 60% of the total systematic variation in the change-score distribution;
- Mimics regression-to-the-mean not via random error but through deterministic geometry.

Indeed, much of the variation in observed change scores can be parsimoniously accounted for by the structural features of the response scale. Attempts to improve fit with additional parameters yielded negligible gains, suggesting that added model complexity is neither necessary nor theoretically justified in this context.

While the generative PLRC model offers a compelling explanation for scale-induced asymmetries, we now seek to formalize these dynamics using a linear regression framework. This allows us to estimate the effects of upward and downward deviation given a central "corridor" on observed change scores.

Specifically, we define a piecewise linear regression model of the form:

$$\Delta T = \alpha + \beta_L d_L + \beta_U d_U + \varepsilon$$

where α represents the expected change for respondents whose baseline scores lie within the neutral corridor $[2.5, 4.5]$. The variables $d_L = \max(0, L - T^1)$ and $d_U = \max(0, T^1 - U)$ represent directional deviations below and above the corridor bounds $L = 2.5$ and $U = 4.5$, respectively. The coefficients β_L and β_U capture the marginal effects of lower- and upper-deviation on the magnitude and direction of response change.

Table 22: deviation-Only PLRC Model Estimates (All Health Items Combined)

Variable	**Coefficient**	**Std. Error**	**t-value**	**95% CI**
Intercept (`const`)	-0.3080^{***}	0.026	-12.02	$[-0.358, -0.258]$
`dL` (Lower deviation)	0.6487^{***}	0.041	15.78	[0.568, 0.729]
`dU` (Upper deviation)	-0.5166^{***}	0.021	-24.25	$[-0.558, -0.475]$

Model fit statistics:

$N = 6115$, $R^2 = 0.166$, Adjusted $R^2 = 0.166$, F-statistic = 610.4, $p < .001$

AIC = 21310, BIC = 21340, Durbin–Watson = 1.956

Residual diagnostics:

Omnibus test = 40.58 ($p < .001$), Jarque–Bera = 55.93 ($p < .001$)

Skew = -0.077, Kurtosis = 3.443 *Note:* $^{***}p < .001$

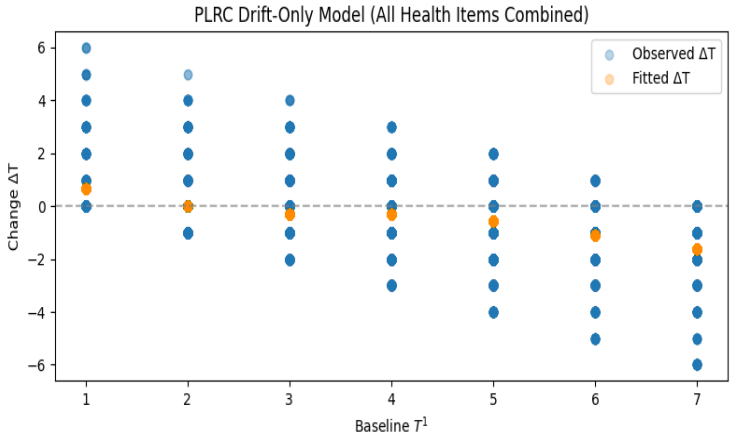


Figure 12: Empirical data (blue)–Health, and PLRC (orange).

The regression results in Table 22 offer empirical confirmation of the PLRC model's core prediction: observed within-person changes are systematically shaped by baseline positions relative to the interior of the scale. Specifically, upward deviation from the lower bound (d_L) is associated with positive change ($\beta_L = 0.6487$), while downward deviation from the upper bound (d_U) predicts negative change ($\beta_U = -0.5166$). These effects are both large and statistically significant ($p < .001$), and their opposing signs capture the bidirectional "reflective" pressure hypothesized by the corridor

model.

The intercept ($\alpha = -0.3080$) indicates a modest negative trend even for respondents whose baseline scores fall within the neutral corridor ($[2.5, 4.5]$). This aligns with prior results showing global downward drift in the empirical distribution. Together, these coefficients suggest that much of the observed variation in change can be explained without invoking treatment effects, psychological shifts, or random noise: the bounded geometry of the scale and respondents' positions on it suffice to generate directional change.

Model diagnostics support this interpretation. The model explains roughly 17% of the variance in change scores ($R^2 = 0.166$)—a modest achievement for a structural model with only three predictors and no covariates. Residuals exhibit slight negative skew, but standard tests suggest no major violations of normality.

Table 23: Correlation Matrix

Variable	T^1	ΔT	d_L	d_U
T^1	1.000			
ΔT	-0.430	1.000		
d_L	-0.724	0.294	1.000	
d_U	0.843	-0.364	-0.329	1.000

Despite the absence of any exogenous manipulation or treatment (i.e., no between-wave stimulus), the structural associations among baseline scores T^1, observed changes ΔT, and the deviation terms d_L and d_U remain robust. As shown in Table 22, the deviation terms are strongly correlated with initial position: d_L is negatively correlated with T^1 ($r = -0.724$), while d_U is positively correlated ($r = 0.843$). These patterns follow directly from their definitions:

$$d_L = \max(0, L - T^1), \quad d_U = \max(0, T^1 - U),$$

such that higher initial scores reduce upward deviation potential (d_L) and increase downward deviation (d_U).

Both deviation terms also exhibit directionally consistent and meaningful correlations with observed change. Specifically, upward deviation (d_L) correlates positively with ΔT ($r = 0.294$), while downward deviation (d_U) correlates negatively ($r = -0.364$). This supports the central claim of the PLRC framework: even in the absence of external input, respondents tend to shift away from the scale's boundaries and toward the "open space," generating asymmetric, RTM-like patterns that arise from the interaction between scale geometry and response processes, rather than from random fluctuation. We turn now turn to a third and final two-wave survey.

4. Bounded Belief

4.1 Data and Source (Anglin, 2019)

This secondary analysis draws on one of three studies in which participants were presented with evidence related to religious beliefs. The selected study, conducted via Amazon Mechanical Turk, investigated how different types of evidence affect belief revision. Participants (N = 427; 160 men, 264 women, 3 transgender; M_{age}=35.85, SD=12.90) were randomly assigned to one of two evidence conditions: religion-enhancing ($n = 213$) or religion-disparaging ($n = 214$). Each group reviewed five research summaries, collectively describing ten empirical studies.

The experimental design included a within-subjects factor assessing beliefs at three time points: Time 1 (pre-evidence), Time 2 (post-evidence), and a retrospective rating of perceived Time 1 beliefs. At each time point, participants rated the relationship between religiosity and life outcomes on a 9-point scale (1 = strongly negative, 9 = strongly positive). Religious affiliations were diverse, with the majority identifying as Christian (n = 220), followed by Agnostic (75), Atheist (60), and others. This analysis is conducted independently of the original authors' interpretations (see Figure 13 for source).

4.2 Conditions

In Figure 13, we observe the proportion of change directions (y-axis) given each initial value (x-axis), T^1 on [1, 9]. Within the religion-disparaging condition (0, blue), change became disproportionately negative as initial beliefs approached the upper bound. In contrast, the religion-enhancing condition (1, red) became disproportionately positive as the initial beliefs approached the lower bound.

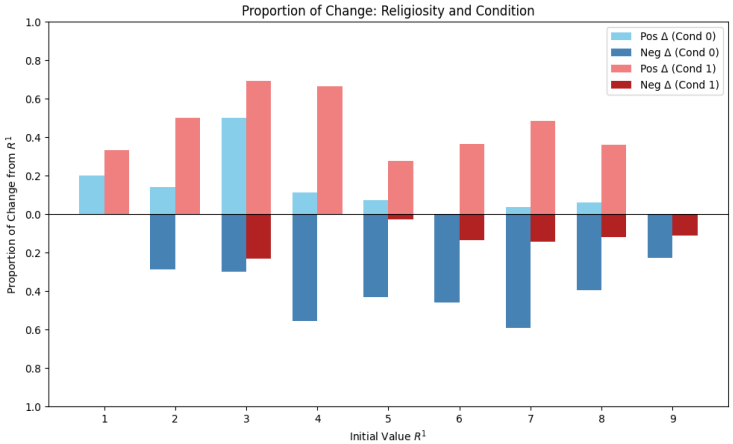


Figure 13: Anglin, 2019

Consistent with the approach in Section 2, we stratify participants based on their initial beliefs relative to the scale midpoint, distinguishing between those above and those at or below the midpoint. Unlike the uniform evidence condition employed by Vlasceanu et al., however, the present study includes an experimental manipulation: participants were randomly assigned to receive either religion-enhancing (C = 1) or religion-disparaging (C = 0) information between measurement waves.

Table 24: Proportion of Change Outcomes by Belief and Condition Groups

Belief Group	**Condition**	**Positive Change**	**Negative Change**	**No Change**
$T^1 \leq 5$	$C = 0$	0.13	0.40	0.47
$T^1 \leq 5$	$C = 1$	0.37	0.05	0.58
$T^1 > 5$	$C = 0$	0.03	0.43	0.55
$T^1 > 5$	$C = 1$	0.31	0.13	0.56

At first glance, the dataset appears to exhibit expected directional effects: negative belief change is disproportionately higher in the religion-disparaging condition (C = 0), while positive belief change is disproportionately higher in the religion-enhancing condition (C = 1). This alignment with the experimental manipulation suggests face-validity of the intervention effects.

To assess this more rigorously, however, we compare the proportion of incongruent belief changes relative to all observed changes, stratified by experimental condition and initial belief group (T^1).

$$\Pr(\Delta^+ \mid C = 0, T^1 \leq 5) = \frac{0.13}{0.13 + 0.40} \approx 0.25$$
$$\Pr(\Delta^- \mid C = 1, T^1 \leq 5) = \frac{0.05}{0.05 + 0.37} \approx 0.12$$

$$\Pr(\Delta^+ \mid C = 0, T^1 > 5) = \frac{0.03}{0.03 + 0.43} \approx 0.07$$
$$\Pr(\Delta^- \mid C = 1, T^1 > 5) = \frac{0.13}{0.13 + 0.31} \approx 0.30$$

These proportions mirror the trend noted in Section 2 regarding differences in consistency with the evidence. For $T^1 \leq 5$, participants in group $C = 0$ had a higher inconsistency rate (25%) than those in $C = 1$ (12%). This indicates that *open movement* against the evidence was about twice as

frequent in the disparaging group. In other words, participants whose initial positions fell below the midpoint were more likely to misinterpret the evidence in the direction corresponding to the side with a greater density of support values, rather than toward the side with a smaller density.

For $T^1 > 5$, participants in group $C = 1$ showed greater inconsistency (30%) than those in $C = 0$ (7%), suggesting again that open movement *against the evidence* was roughly *4x more likely* in the enhancing group. Once again, participants whose initial positions fell above the midpoint were more likely to misinterpret the evidence in the direction corresponding to the side with a greater density of support values, rather than toward the side with a smaller density.

Given the symmetrical division of conditions, this pattern of "open movement bias" also extends to the proportion of changes that are congruent with the condition, as it naturally complements the rate of incongruent changes within each subgroup and condition.

$$\Pr(\Delta^- \mid C = 0, T^1 \leq 5) = \frac{0.4}{0.13 + 0.40} \approx 0.75$$
$$\Pr(\Delta^+ \mid C = 1, T^1 \leq 5) = \frac{0.37}{0.05 + 0.37} \approx 0.88$$

$$\Pr(\Delta^- \mid C = 0, T^1 > 5) = \frac{0.43}{0.03 + 0.43} \approx 0.93$$
$$\Pr(\Delta^+ \mid C = 1, T^1 > 5) = \frac{0.31}{0.13 + 0.31} \approx 0.70$$

Now, congruent changes are more frequent in the direction corresponding to the side with a greater density of support values, rather than toward the side with a smaller density.

We again conduct a permutation test to assess whether the observed cross-bin variance in response change exceeds that obtained in 5,000 permutations, where $T^1 \in [1, 9]$. The cross-bin variance is defined as

$$\mathrm{Var}(\bar{\Delta}) = \frac{1}{9} \sum_{r=1}^{9} \left(\bar{\Delta}_r - \bar{\bar{\Delta}} \right)^2,$$

where $\bar{\Delta}_r$ is the mean response change for bin r, and $\bar{\bar{\Delta}}$ is the grand mean across all bins.

To generate the null distribution, we permute the T^1 labels and recompute the variance 5,000 times. The null hypothesis is that the observed variance arises by chance.

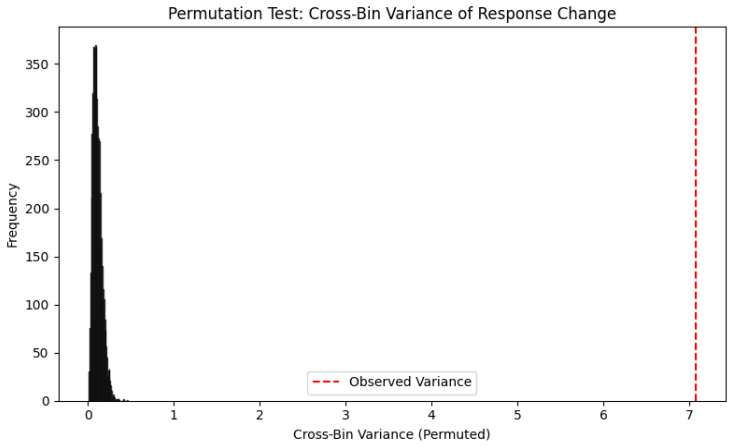


Figure 14: Permutation test for cross-bin variance in response change.

The permutation test offers strong statistical evidence that the magnitude of response change, defined as $\Delta T = T^2 - T^1$, varies systematically across baseline bins $T^1 \in \{1,\ldots,9\}$. The observed cross-bin differences in mean change are unlikely to arise under the null hypothesis of random baseline assignment.

Formally, the test evaluates

$$H_0 : \bar{\Delta}_1 = \bar{\Delta}_2 = \cdots = \bar{\Delta}_9,$$

which asserts that mean response change is uniform across bins. The observed variance,

$$\mathrm{Var}(\bar{\Delta})_{\mathrm{obs}} = 7.07,$$

far exceeds the distribution of variances generated under 5,000 random permutations of T_1, where values typically range from 0.1 to 0.2 and seldom exceed 1.0. The resulting empirical p-value is

$$p < \frac{1}{5000} = 0.0002 \quad \text{(reported as 0.00000)},$$

indicating that none of the permuted datasets produced a variance as large as the observed.

These findings suggest that, once again, response change exhibits bin-dependent heterogeneity. Models that assume constant change across the scale (e.g., OLS) may therefore fail to capture relevant structural variation. In contrast, bounded models such as the PLRC, which explicitly incorporate asymmetry and floor/ceiling constraints, may better reflect the empirical patterns observed in the data.

The estimated regression model is:

$$\Delta T = \beta_s(s - T^1) + \beta_L d_L + \beta_U d_U + \varepsilon$$

where T^1 is baseline belief score, T^2 is follow-up belief score, $s = 1$ if condition = 0 (religion disparaging) and 9 if condition = 1(religion enhancing), $d_L = \max(0, L - T_1)$, and $d_U = \max(0, T_1 - U)$.

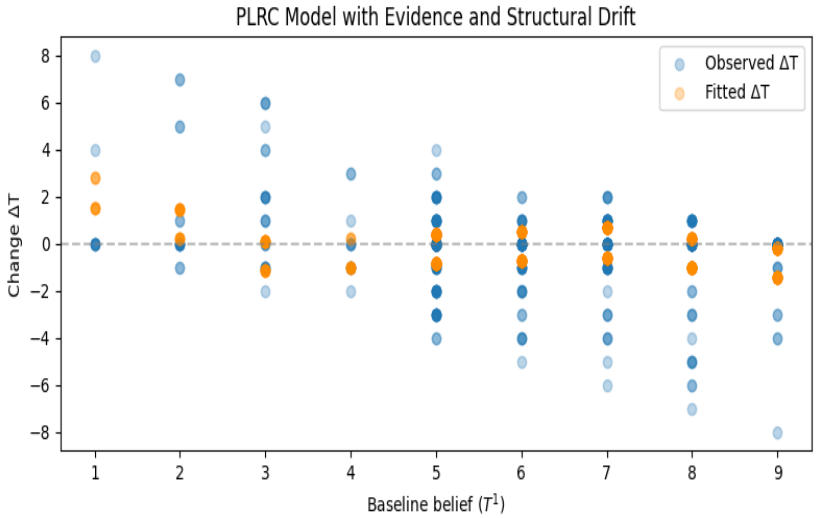


Figure 15: Observed and model-predicted changes in belief across baseline bins. Each point represents an individual or model-generated response change $\Delta T = T^2 - T^1$, plotted against baseline belief $T^1 \in \{1, \ldots, 9\}$. Blue points indicate observed changes, while orange points show fitted values from the PLRC model with evidence and structural deviation.

Table 25: OLS Regression Estimates for Normalized Surprise and deviation Components

Variable	**Coefficient**	**Std. Error**	**t-value**	**95% CI**
Intercept	-0.3055^{**}	0.096	-3.18	$[-0.494, -0.117]$
$(s - T^1)$	-0.1382^{***}	0.018	-7.85	$[-0.173, -0.104]$
d_L	1.4760^{***}	0.251	5.88	[0.982, 1.970]
d_U	-0.5536^{***}	0.125	-4.43	$[-0.799, -0.308]$

Model fit statistics:

$N = 427$, $R^2 = 0.175$, Adjusted $R^2 = 0.169$, F-statistic = 29.96, $p < .001$

AIC = 1653, BIC = 1669, Durbin–Watson = 1.931

Residual diagnostics:

Omnibus test = 35.97 ($p < .001$), Jarque–Bera = 150.50 ($p < .001$)

Skew = -0.165, Kurtosis = 5.890 *Note:* $^{**}p < .01$, $^{***}p < .001$

Table 26: Correlation Matrix

Variable	ΔT	$(s - T^1)$	d_L	d_U
ΔT	1.000	0.481	0.212	-0.126
$(s - T^1)$	0.481	1.000	0.210	-0.278
d_L	0.212	0.210	1.000	-0.126
d_U	-0.126	-0.278	-0.126	1.000

Similar to Section 1 analysis, this model reflects a covariance structure where ΔT(the change in belief) is partially explained by the difference between the target value s and the baseline belief T^1, as well as the deviation components d_L and d_U. As the value of s shifts between 1 and 9 depending on the condition, the linear structure of the model remains stable, ΔT is still influenced by how far T^1 is from the bounds L and U, through d_L and d_U. The correlation matrix shows that $(s - T^1)$ has a moderate positive correlation with $\Delta T (r = 0.481)$ and shares similar but weaker relationships with

d_L and d_U, which have low-magnitude correlations with each other and with ΔT. This consistency in structure once again suggests that even as the directional term $(s - T^1)$ varies due to changes in s, the roles of deviation terms d_L and d_U remain functionally invariant, maintaining a stable covariance relationship within the model and evidence discrepancy term.

5. Limitations and Considerations

5.1 Latent Variable Modeling

We evaluated several latent variable modeling frameworks for estimating unobserved constructs such as political motivation, religiosity, and fear. However, we chose not to apply latent variable models, as the analysis was secondary and the constructs of interest were not well enough understood to warrant latent modeling. Our understanding of these constructs raises definitional and measurement ambiguities, which undermine interpretability in the absence of a validated measurement framework. Additionally, latent model outputs are often highly sensitive to specification choices (e.g., priors, constraints, link functions), introducing model-dependent variability that complicates inference.

That said, we do implement the PLRC within a Bayesian model in PyMC, using data from Anglin (2019). The PLRC adjusts unconstrained latent change (δ_{raw}) to remain within the bounded response range (1–9), while softly encouraging values to remain within a central corridor defined by $[L = 3, U = 7]$. Posterior inference exhibited no sampling divergences, and posterior predictive checks indicated stable numerical performance. We conclude that the PLRC offers a computationally tractable, differentiable, and interpretable approach to modeling bounded belief change within a Bayesian framework.

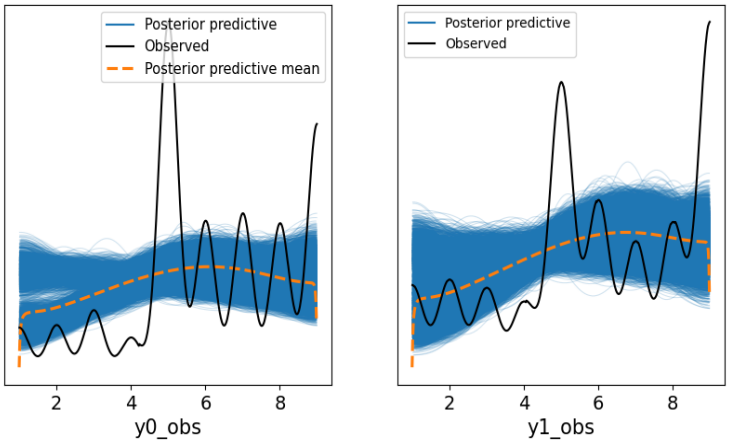


Figure 16: Posterior predictive checks for observed responses at Time 1 (y_0) and Time 2 (y_1). Black lines indicate the empirical distribution of observed Likert-scale responses, while blue lines represent posterior predictive draws from the model. The dashed orange line shows the posterior predictive mean. The model captures general trends but underestimates local spikes in the response distribution, particularly in upper-scale modes.

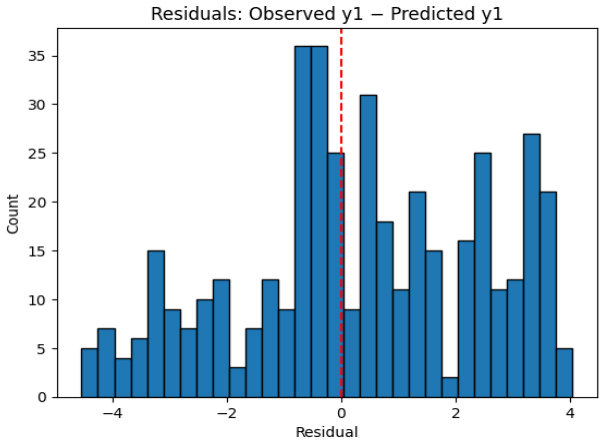


Figure 17: Histogram of residuals for Time 2 predictions (Observed y_1 minus Predicted y_1). The residual distribution is approximately centered around zero (dashed red line), suggesting the model is generally unbiased. However, the presence of heavy tails and asymmetry indicates that the model underfits more extreme responses, particularly on the upper end of the scale. The RMSE is 2.16, which is relatively large given the bounded response range of [1, 9].

5.2 Dichotomizing Variables

Scholars have long cautioned against dichotomizing continuous variables, noting that it can compromise validity and obscure meaningful effects by imposing arbitrary thresholds (Greenland, 2017; Streiner, 2002). As Streiner argues, "(1) results can become 'useless' if the chosen cutpoint is

altered, and (2) effects that are statistically significant in continuous form may lose significance when reduced to binary categories."

With respect to the first concern, our findings suggest the opposite: the observed response patterns are robust to reasonable variation in the cutpoint. Altering the threshold does not undermine the central conclusions. Rather, the overall structure of response change remains intact—respondents starting near the upper bound tend to show larger negative shifts, while those near the lower bound exhibit stronger positive changes. In contrast, individuals near the midpoint show smaller and more variable changes, both in magnitude and direction. This reflects a meaningful gradient in the data, not an artifact of dichotomization.

Regarding the second concern, our inferences do not rely solely on statistical significance. Instead, dichotomization is used to clarify theoretically motivated contrasts between distinct response states (e.g., high vs. low initial beliefs), thereby sharpening interpretability. In this context, binning enhances the precision of inferential targets rather than oversimplifying the data.

Nonetheless, we acknowledge that dichotomizing continuous variables entails trade-offs. It reduces within-group variance and may obscure nonlinear associations, particularly near the midpoint. Although our results are robust to different cutpoints, dichotomization inevitably imposes discretization costs that should be weighed against the benefits of analytic clarity.

Why $\hat{\rho}$?

Raw change scores (ΔT) are widely used in longitudinal research across psychology, education, and health sciences. However, such measures do not account for the inherent asymmetry in the relationship between scale bounds and initial response. Respondents situated near the lower or upper limits of the scale have less room to exhibit further change, introducing systematic bias when comparing individuals with different baseline values. While standardized response means (SRMs) and effect size metrics such as Cohen's d offer adjustments for variability, they do not explicitly address these constraints.

A longstanding methodological literature has cautioned that change scores are not linearly

interpretable, due to statistical dependencies and biases such as measurement attenuation (Cronbach & Furby, 1970) and regression to the mean (Barnett, et al., 2005). These works help frame the interpretive challenges introduced by survey scales—particularly in the use of change scores—but they do not address boundary-induced asymmetries that arise in multi-wave data.

The $\hat{\rho}$ index serves as a diagnostic visualization tool and, in its leave-one-out form ($\hat{\rho}^{\text{loo}}$), as a potential predictor for hypothesis testing. While $\hat{\rho}^{\text{loo}}$ mitigates endogeneity concerns, it may underrepresent individual-level variability, potentially obscuring idiosyncratic response behavior.

The index assumes symmetric latent change potential across the scale, an assumption that may not hold empirically due to cognitive bias. Although its formula—change divided by available directional space—is intuitively appealing, it lacks a psychometric foundation, raising concerns about ratio-scale assumptions applied to ordinal data.

As a derived score rather than an estimated parameter, $\hat{\rho}$ is not easily embedded in standard inferential frameworks limiting its integration with multivariate or multi-scale analyses

5.3 PLRC & Extensions

By attributing asymmetries in observed change scores to structural features of the response scale, the PLRC framework offers a compelling account of scale-induced distortions. However, this explanatory approach carries the risk of obscuring meaningful psychological processes. In particular, asymmetric patterns of change—such as greater resistance to belief updating at one end of the scale—may reflect substantive individual differences in responsiveness rooted in mechanisms like motivated reasoning, selective exposure, or cognitive dissonance. In politically charged or clinical contexts, these asymmetries may themselves be theoretically significant. Over-attributing such effects to measurement artifacts may inadvertently diminish construct validity by treating psychologically meaningful variation as noise.

While the PLRC model represents an attempt at an important conceptual advance by formalizing how bounded scale geometry can induce apparent change, it is currently implemented as a stand-

alone deterministic transformation, outside of a fully integrated estimation framework. Embedding the logic of the PLRC within latent variable models—such as Bayesian latent change frameworks, item response theory models with bounded traits, or dynamic response process models—offers a promising path forward. Such integration would allow researchers to assess whether the PLRC's implied change distributions recover true latent dynamics under known constraints, and to evaluate model identifiability and parameter recovery within a principled inferential architecture. It would also facilitate stronger links to measurement theory and construct validity by jointly modeling both the structural and psychological determinants of response behavior.

Finally, an open question concerns the origin of the reflective corridor itself. In the present framework, the corridor is specified *a priori* to capture hypothesized constraints on change. However, it remains possible that such a corridor emerges empirically from the interaction of boundedness, reporting behavior, and latent tendencies toward moderation or anchoring. Investigating whether such corridors arise organically, and under what theoretical or empirical conditions, constitutes a valuable direction for future research.

6. Conclusion

All survey methods contend with expressive and epistemic limitations. Respondents are often required to articulate judgments regardless of prior reflection, resulting in scale-equivalent responses from individuals with substantively different levels of engagement or conviction. Open-ended formats, while less constraining, rely on variable linguistic and cognitive capacities. As a result, the medium of response—whether numerical, verbal, or behavioral—may not merely transmit but shape what is expressed.

This raises a fundamental question: to what extent is belief—or the degree of belief—constituted by the structure of its articulation? If the format influences the nature of the response, then measurement tools are not neutral vessels but active participants in the construction of self-reported attitudes.

We identify three general mechanisms by which the expression of a latent construct may be systematically misrepresented. These differ in locus—external vs. internal—and in whether the medium mediates comprehension, expression, or transformation:

- **Epistemic Opacity (External Comprehension Through Medium):** The researcher's access to a respondent's latent state is mediated by the structure of the measurement instrument. Features of the response format—such as scale geometry, verbal framing, or response options—can introduce systematic distortions between the underlying state and its integrated representation.
- **Self-Opacity (Internal Expression Through Medium):** Respondents may lack full introspective access to their own beliefs, preferences, or affective states. Alternatively, they may struggle to map these internal states onto the medium of expression or provided response options, resulting in imperfect or inconsistent self-reporting.
- **Construct Plasticity (Internal Alteration By Medium):** The act of responding may not merely express a pre-existing state, but actively reshape it. In this case, the measurement process participates in the construction of the realized state itself, such that cognition and expression co-construct response through interaction with the response medium.

This framework clarifies a central challenge in the measurement of psychological dispositions: the medium of expression is not neutral. Each of the three mechanisms—epistemic opacity, self-opacity, and construct plasticity—represents a distinct way in which measurement instruments may shape, obscure, or even generate the very dispositions they aim to assess. Dispositions, by definition, are latent. But as this taxonomy reveals, the process of eliciting these latent states can interfere with their interpretation: external opacity limits the researcher's comprehension of the state, internal opacity limits the subject's ability to express it, and construct plasticity implies that the act of measurement and expression may alter the state itself.

Together, these dynamics raise critical concerns for construct validity, comparability across instruments, and the interpretation of individual differences. They also motivate a broader research

agenda: to develop measurement strategies that are not only psychometrically reliable but also epistemologically aware—capable of accounting for the interpretive and generative roles that response formats play in the measurement of latent constructs.

Final Remarks

Our results are drawn from three two-wave panel datasets. While robust within these contexts, generalizing the observed structural effects to other domains requires further empirical testing. Future research should examine applicability to non-unidimensional constructs, multi-wave designs, and alternative response formats, including all non-binary scales. Experimental manipulations of scale structure offer a promising strategy for isolating the causal influence of geometric constraints on observed change.

A central aim for future research is to more precisely disentangle structural artifacts from genuine psychological asymmetries in response dynamics. Although our findings demonstrate that directional asymmetries are fairly robust across two-wave panel datasets, the origins of these effects remain uncertain—whether they arise from respondent cognition, the structure of the measurement instrument, or an interaction between the two. Beyond theoretical clarification, we also seek to enhance the applied value of the PLRC framework by investigating whether, and how, evidence influences response patterns in polls and other real-world settings.

We recognize skepticism: "Do survey boundaries matter this much?" "Isn't this just regression to the mean?" These questions merit serious consideration. But our findings suggest something more fundamental. The observed asymmetries are not mere statistical artifacts but structural distortions emerging from how bounded scales shape respondent behavior. First, proximity to scale boundaries systematically conditions both the direction and magnitude of response change. This is not trivial. In some cases, over one-third of updates became misaligned with the evidentiary direction, and half of those flawed changes were unmotivated. We contend that such misalignment can largely be attributed to initial position proximity to boundaries alone. Our PLRC model and the diagnostic $\hat{\rho}$

index show that available "headroom" is not a nuisance term—it is better understood as a geometric force.

Second, these effects cannot be reduced to classical regression toward the mean. RTM assumes symmetric reversion around a central tendency; what we observe is position-induced, directional bias that persists even after normalization. LOESS modeling reveals inflection points that RTM cannot explain, and which require a structural rather than stochastic account.

This is not correction of measurement error—it is diagnosis of measurement structure, more precisely, its influence on response behavior. The instrument does not simply reflect beliefs; it interacts with them, shaping their expression through the constraints and affordances of its design.

We extend long-standing insights about framing and question wording to include the geometry of the response scale itself. This is not a rejection of standard approaches, but a refinement: a shift from viewing bias as noise to understanding it as structure-induced behavior. Our framework draws on ecological psychology, Gibson's (1975) notion of affordances, the difference being, in our case, the response scale constitutes the environment, and its geometry defines the "action space" afforded to respondents.

Respondents may not treat "room to move" on a response scale as a passive structural feature, but as an affordance—implicitly signaling the direction or magnitude of acceptable change. From this perspective, response scale geometry may function not only as a constraint, but as an active cue influencing behavior.

The PLRC model operationalizes this by capturing how boundedness and spatial asymmetry affect observed response dynamics. We suggest that, across the scale, respondents experience a form of directional "pressure" induced by proximity to the bounds.

Accordingly, observed asymmetries in change scores may reflect not artifacts, but patterns of engagement with the response instrument. Disentangling scale-induced structure from psychologically meaningful change will require experimental manipulations of response geometry and broadened empirical research, combined with formal modeling frameworks that incorporate both cognitive and

structural mechanisms.

In sum, rather than correcting bias post hoc, we advocate diagnosing and modeling the structural conditions that generate it.

References

Amaya, A. E., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology, 8*(1), 89–119. https://doi.org/10.1093/jssam/smz056

Anglin, S. M. (2019). Do beliefs yield to evidence? Examining belief perseverance vs. change in response to congruent empirical findings. *Journal of Experimental Social Psychology, 82*, 176–199. https://doi.org/10.1016/j.jesp.2019.02.004

Balyan, P., & Sanjeev, M. (2014). Response order effects in online surveys: An empirical investigation. *International Journal of Online Marketing, 4*(2), 28–44. https://doi.org/10.4018/ijom.2014040103

Batchelor, J. H., & Miao, C. (2016). Extreme response style: A meta-analysis. Journal of Organizational Psychology, 16(2), 51–62

Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford University Press.

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. International Journal of Epidemiology, 34(1), 215–220. https://doi.org/10.1093/ije/dyh299

Bartholomew, D. J. (1996). *The statistical approach to social measurement* (1st ed.). Academic Press.

Bartholomew, D. J., Steele, F., & Moustaki, I. (2008). *Analysis of multivariate social science data* (2nd ed.). CRC Press.

Borghans, L., Duckworth, A. L., Heckman, J., & ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources, 43*(4), 972–1059. https://doi.org/10.1353/jhr.2008.0017

Bourdieu, P. (1992). *The logic of practice* (R. Nice, Trans.). Stanford University Press. (Original work published 1980)

Celhay, P. A., Meyer, B. D., & Mittag, N. (2022, January). What leads to measurement errors? Evidence from reports of program participation in three surveys (NBER Working Paper No. 29652). National Bureau of Economic Research. https://doi.org/10.3386/w29652

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"—or should we? Psychological Bulletin, 74(1), 68–80. https://doi.org/10.1037/h0029382

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer. https://doi.org/10.1007/978-1-4757-3990-9

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland, 15*, 246–263. https://doi.org/10.2307/2841583

Gibson, J. J. (2014). *The ecological approach to visual perception: Classic edition* (1st ed.). Psychology Press. https://doi.org/10.4324/9781315740218 (Originally published in 1979.)

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*(1), 1–19. https://doi.org/10.1017/S0140525X97000010

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European journal of epidemiology, 31(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.

Haaland, I. K., Roth, C., Stantcheva, S., & Wohlfart, J. (2025, April). Understanding economic behavior using open-ended survey data [Manuscript submitted for publication]. *Journal of Economic Literature*.

Höhne, J. K., Lenzner, T., & Claassen, J. (2025). Automatic speech-to-text transcription: Evidence from a smartphone survey with voice answers. *International Journal of Social Research Methodology*. Advance online publication. https://doi.org/10.1080/13645579.2024.2443633

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology, 96*(5), 933–948. https://doi.org/10.1037/a0014747

Kalton, G., & Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society. Series A (General), 145*(1), 42–73. https://doi.org/10.2307/2981421

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research, 22*(3), 320–342. https://doi.org/10.1093/ijpor/edq001

Lord, E. M. (1956). The measurement of growth. *ETS Research Bulletin Series, 1956*(1), i–22. https://doi.org/10.1002/j.2333-8504.1956.tb00058.x

Lord, E. M. (1967). A paradox in the interpretation of group comparisons. Psychological Bulletin, 68, 304–305. doi:10.1037/h0025105

Mee, R. W., & Chua, T. C. (1991). Regression toward the mean and the paired sample t test. *The American Statistician, 45*(1), 39–42. https://doi.org/10.2307/2685237

Meyer, B. D. (2015). Household surveys in crisis. *Journal of Economic Perspectives, 29*(4), 199–226. https://doi.org/10.1257/jep.29.4.199

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science, 1*(2), 110–122. https://doi.org/10.1111/j.1745-6916.2006.00008.x

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data*

analysis methods (2nd ed.). SAGE Publications.

Rasinski, K. A., Lee, L., & Krishnamurty, P. (2012). Question order effects. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 229–248). American Psychological Association. https://doi.org/10.1037/13619-014

Salganik, M.J. (2018). Bit by Bit: Social Research in the Digital Age. United Kingdom: Princeton University Press.

Samuelson, P. A. (1938). A note on the pure theory of consumers' behaviour. *Economica, 5*(17), 61–71. https://doi.org/10.2307/2548836

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica, 15*(60), 243–253. https://doi.org/10.2307/2549561

Scheuren, F. (2004). *What is a Survey?* United States: American Statistical Association.

Shaw, J. (2014, March–April). Why "big data" is a big deal: Information science promises to change the world. *Harvard Magazine*. Retrieved from https://www.harvardmagazine.com/2014/02/why-big-data-is-a-big-deal

Siegrist, M., & Bearth, A. (2021). Worldviews, trust, and risk perceptions shape public acceptance of COVID-19 public health measures. *Proceedings of the National Academy of Sciences of the United States of America, 118*(24), e2100411118. https://doi.org/10.1073/pnas.2100411118

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118. https://doi.org/10.2307/1884852

Singleton, R. A., Jr., & Straits, B. C. (2009). *Approaches to social research* (5th ed.). Oxford University Press.

Smaldino, P. E., & McElreath, R. (2016, September 21). The natural selection of bad science. *Royal Society Open Science, 3*(9), Article 160384. https://doi.org/10.1098/rsos.160384

Speer, D. C. (1999). What is the role of two-wave designs in clinical research? Comment on Hageman and Arrindell. *Behaviour Research and Therapy, 37*(12), 1203–133. https://doi.org/10.1016/s0005-7967(99)00034-0

Stantcheva, S. (2022, September). How to run surveys: A guide to creating your own identifying variation and revealing the invisible (NBER Working Paper No.30527). National Bureau of Economic Research. https://doi.org/10.3386/w30527

Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research, 6*(2), 103–114. https://doi.org/10.1177/096228029700600202

Streiner, D. L. (2002). Breaking up is hard to do: The heartbreak of dichotomizing continuous data. The Canadian Journal of Psychiatry, 47(3), 262–266. https://doi.org/10.1177/070674370204700307

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. Jossey-Bass.

Tourangeau, R. (1984). Cognitive science and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). National Academy Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). The psychology of survey response. Cambridge University Press.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology, 4*(3), 173–187. https://doi.org/10.1207/s15326969eco0403_3

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vlasceanu, M., Morais, M. J., & Coman, A. (2021). The effect of prediction error on belief update across the political spectrum. *Psychological Science, 32*(6), 916–933. https://doi.org/10.1177/0956797621995208

Vogt, W. P. (2005). Scale attenuation effect. SAGE Research Methods. https://doi.org/10.4135/9781412983907 (Original work published in the SAGE Encyclopedia of Social Science Research Methods).

Warren, W. H., Jr. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 683–703. https://doi.org/10.1037/0096-1523.10.5.683

Wells, A. J. (2002). Gibson's Affordances and Turing's Theory of Computation. *Ecological Psychology, 14*(3), 140–180. https://doi.org/10.1207/S15326969ECO1403_3

Wooldridge, J. M. (2005). *Introductory econometrics: A modern approach* (3rd ed.). South-Western College Publishing.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Yule, G. U. (1932). Why do we sometimes get nonsense-correlations between time-series?—A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society, 95*(1), 1–63. https://doi.org/10.2307/2341482

Appendix

Original Research Designs

(1) Original Research Vlasceanu *et al.*, 2021

The researchers assign participants to either a prediction group or control group. Participants in the prediction group first assessed the accuracy of each proposition (Wave-1 evaluation), then predicted the correct fact from a set of 12 plausible options, then received the actual fact, and subsequently reassessed their initial judgment (Wave-2 evaluation). In contrast, control group participants assessed the proposition (Wave-1), received the fact, and then reassessed their initial judgment (Wave-2) without making any prediction. The primary findings indicate that the magnitude of prediction errors served as a linear predictor of response updating, with larger errors prompting more substantial response revision than cases without prediction engagement (control group). Importantly, these effects were consistent across both Democrats and Republicans and were observed

consistently across Democratic-leaning, Republican-leaning, and neutral propositions.

Supplementary Materials: Statements

	Type	Statement
1	Neutral	The average income of US government employees is modest.
2	Neutral	Very few Americans identify as vegetarian.
3	Neutral	A small number of Americans die of electrocution by their toasters every year.
4	Neutral	A large proportion of American households don't save anything from their income.
5	Neutral	Diabetes has huge costs for the US economy.
6	Neutral	Pneumonia is responsible for many child deaths.
7	Neutral	Many American adults exercise on a daily basis.
8	Neutral	Left and right handed people earn equivalent incomes.
9	Neutral	CPR is an effective life-saving method.
10	Neutral	Shark attack rates are similar for men and women.
11	Neutral	Buses and cars are just as likely to be involved in road accidents.
12	Neutral	Only a modest number of Twitter accounts are fake.
13	Democratic	The US has loose gun laws.
14	Democratic	The US government spends little for climate related research.
15	Democratic	Obamacare has successfully decreased the number of uninsured Americans.
16	Democratic	Embryonic stem cell therapy is a successful modern treatment method.
17	Democratic	Colleges and Universities are having a positive effect on young generations' futures.
18	Democratic	The Affordable Care Act saved the US a huge amount of money.
19	Democratic	All cities in the US experience more extremely hot days compared to 50 years ago.
20	Democratic	Children in the US are at high risk of witnessing gun violence.
21	Democratic	The US allocates too much of the spending budget to Defense and Military.
22	Democratic	Children raised by same-sex parents are just as adjusted as children raised by opposite-sex parents.
23	Democratic	The amount government assistance to poor families in the US is not high enough.
24	Democratic	Immigrant households in the US rarely access welfare programs.
25	Republican	The US is at great risk of illegal drug activity.
26	Republican	African American women get more abortions than Caucasian women.
27	Republican	Police use of force in the US is not causing that many deaths.
28	Republican	A large proportion of immigrants in the US are not in the workforce.
29	Republican	The amount of US corporate income taxes paid yearly is high.
30	Republican	A large number of undocumented workers are working illegally in the US.
31	Republican	Currently, foreign-born terrorists are a big threat to Americans in the US.
32	Republican	A large percentage of abortions in the US are paid for with public funds.
33	Republican	In the US, men and women are, on average, paid equally for the same job.
34	Republican	The US justice system is fair to racial minorities.
35	Republican	Government regulations have large costs for the US economy.
36	Republican	Small businesses owned by immigrants in the US do not provide that many jobs.

Figure 18: Original Propositions from Vlasceanu et al., 2021

Supplementary Materials: Evidence

	Evidence	Correct answer	Fact
1	Support	Low	The average yearly salary of US government workers is $51 thousand.
2	Support	Low	5% of Americans are vegetarian.
3	Support	Low	300 Americans die every year by electrocution by their toasters.
4	Support	High	47% of American households don't save anything from their income.
5	Support	High	Diabetes costs the US $266 billion annually.
6	Support	High	Pneumonia is responsible for 1 million child deaths worldwide every year.
7	Against	Low	5% of Americans participate in 30 minutes of physical activity every day.
8	Against	Low	Left handed people earn 10% less than right handed people.
9	Against	Low	2% of people who collapse on the street fully recover from receiving CPR.
10	Against	High	90% of people attacked by sharks are men.
11	Against	High	Car accidents are 60 times more likely than bus accidents.
12	Against	High	Twitter suspended 70 million fake accounts in 2018.
13	Support	Low	In the US, 3 of the 50 states require a permit to purchase a rifle.
14	Support	Low	The US 2017 federal funding for climate research was 0.3% of the annual budget.
15	Support	Low	Obamacare has dropped the number of uninsured Americans from 48 million to 29 million.
16	Support	High	75% of patients treated with stem cell procedures have improved.
17	Support	High	College graduates earn on average $1 million more than high school graduates over their lifetime.
18	Support	High	The Affordable Care Act saved the US $2.3 trillion in costs.
19	Against	Low	Compared to 50 years ago, 73% of cities in the US experience more extremely hot days.
20	Against	Low	4% of children in the US have witnessed a shooting in the past year.
21	Against	Low	The US allocates 20% of the yearly federal budget to the Department of Defense.
22	Against	High	Children raised by same-sex parents are twice more likely to experience emotional problems.
23	Against	High	The costs of poverty assistance programs in the US add up to $1 trillion.
24	Against	High	63% of non-citizen households in the US access welfare programs.
25	Support	Low	Every 20 seconds someone is arrested for a drug offense in the US.
26	Support	Low	Caucasian women get 3.5 times fewer abortions than African American women.
27	Support	Low	Deaths resulted from police use of force occur for every 0.0015% of police-civilian contacts.
28	Support	High	35% of immigrants in the US are not in the workforce.
29	Support	High	The amount of corporate income taxes paid in the U.S. each year is $293 billion.
30	Support	High	8 million undocumented workers are working illegally in the US.
31	Against	Low	On average, 1 American per year is killed in the US by foreign-born terrorists.
32	Against	Low	14% of abortions in the US are paid for with public funds.
33	Against	Low	On average in the US, for every dollar a woman makes, a man makes $1.25.
34	Against	High	An African American is 5 times more likely to be imprisoned compared to a white American for a similar crime.
35	Against	High	Government regulations in the US create benefits are 7 times larger than the value of the costs.
36	Against	High	In the US, small businesses owned by immigrants create 10 million jobs.

Figure 19: Between-wave Fact from Vlasceanu et al., 2021

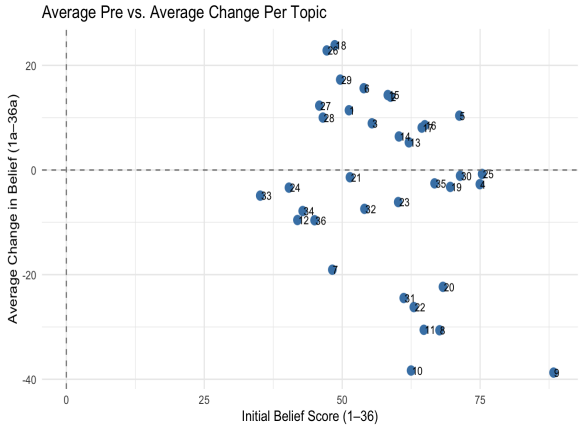


Figure 20: Average Change on Average Initial Response Value by Topic

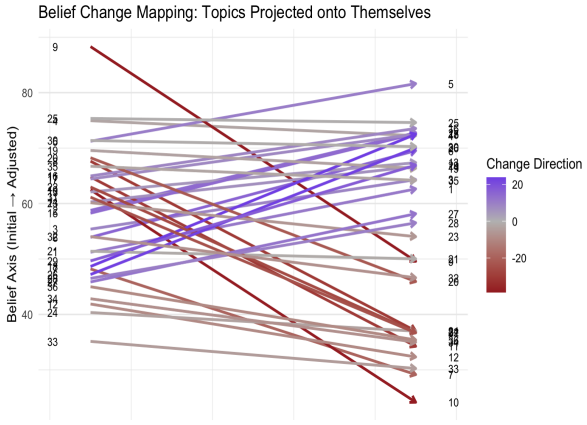


Figure 21: Average Signed Change by Response Topic

Sensitivity & Exploratory Data Analysis Regarding Section 2.

The directional gradient supports the broader hypothesis that scale geometry—not merely evidence type or individual traits—constrains response behavior. Varying bin granularity (e.g., 10 vs. 50 bins) yields qualitatively similar patterns: the magnitude of normalized deviation shifts slightly,

but its sign, slope, and direction remain consistent. This robustness rules out binning artifacts as the source of the observed asymmetry.

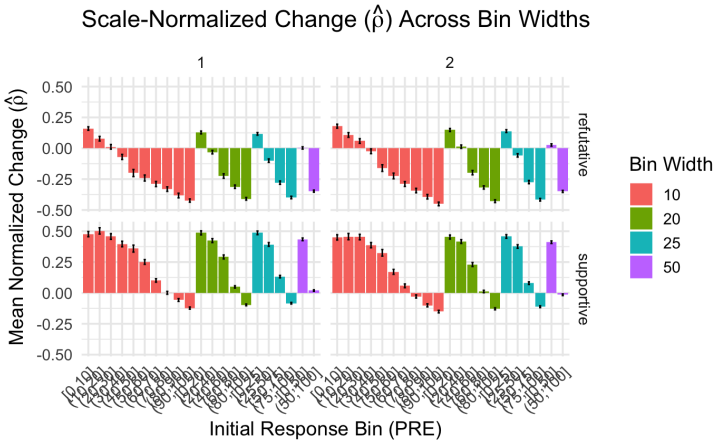
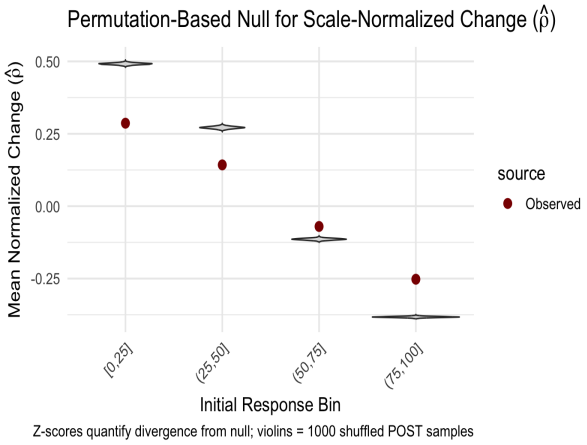
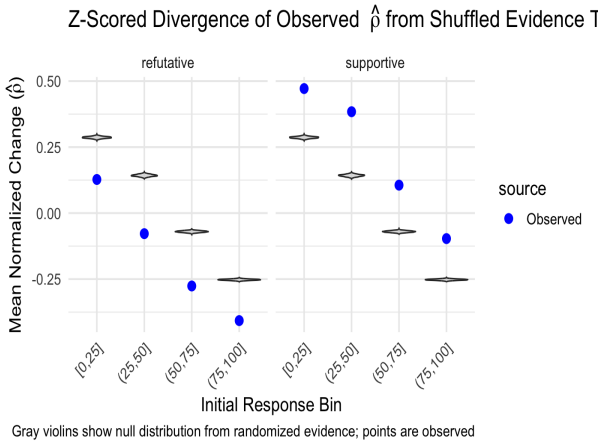


Figure 22: Politics-Study 1

To isolate structural asymmetry from random or treatment-induced variation, we re-estimate $\hat{\rho}(b_0)$ under two randomization procedures: shuffling *POST* values (nullifying individual updating) and permuting *evidence_type* labels (scrambling treatment assignment). In both cases, if the observed $\hat{\rho}$ values lie outside the bulk of the null distribution, the observed asymmetry is unlikely to arise from random updating or treatment confounds. Instead, it reflects a systematic interaction between initial scale position and bounded geometry.



(a) Politics – Study 1



(b) Politics – Study 1

Figure 23: Side-by-side comparison permutations for Study 1.

Table 27: Comparison of Observed vs. Null Mean $\hat{\rho}$ by Initial Bin (Randomizing POST)

PRE Bin	Mean $\hat{\rho}$	Null Mean	Null SD	Z-Score
[0,25]	0.287	0.492	0.0033	-61.91
(25,50]	0.143	0.271	0.0040	-32.23
(50,75]	-0.070	-0.114	0.0031	14.08
(75,100]	-0.252	-0.383	0.0021	63.24

Table 28: Observed vs. Null Mean $\hat{\rho}$ by Initial Bin and Evidence Type

PRE Bin	Evidence Type	Mean $\hat{\rho}$	Null Mean	Null SD	Z-Score
[0,25]	Refutative	0.127	0.287	0.0037	-42.53
	Supportive	0.472	0.287	0.0038	49.38
(25,50]	Refutative	-0.078	0.143	0.0043	-51.55
	Supportive	0.384	0.143	0.0043	56.44
(50,75]	Refutative	-0.276	-0.070	0.0032	-64.09
	Supportive	0.106	-0.070	0.0032	54.79
(75,100]	Refutative	-0.407	-0.252	0.0025	-61.33
	Supportive	-0.097	-0.252	0.0025	61.61

Table 29: Contingency Table of PRE Group by Change Category

PRE Group	Negative	No Change	Positive
PRE > 50	22,389	3,596	13,801
PRE ≤ 50	8,056	1,556	14,574

Chi-Square Test Results

Pearson's Chi-squared Test: $X^2 = 4010.91$, $df = 2$, $p < 2.2 \times 10^{-16}$

Table 30: Direction of movement by party, evidence-type, and starting bin (*rows sum to 1*).

Party	Evidence	PRE bin	Stay	Up	Down
1	refutative	2.5	0.55	0.45	0.00
1	supportive	2.5	0.23	0.77	0.00
2	refutative	2.5	0.53	0.47	0.00
2	supportive	2.5	0.30	0.70	0.00
1	refutative	52.5	0.14	0.28	0.58
1	supportive	52.5	0.15	0.69	0.16
2	refutative	52.5	0.15	0.28	0.56
2	supportive	52.5	0.17	0.61	0.22
1	refutative	97.5	0.28	0.00	0.72
1	supportive	97.5	0.60	0.00	0.40
2	refutative	97.5	0.24	0.00	0.76
2	supportive	97.5	0.53	0.00	0.47

Table 31: Binned Belief Change Summary with Proportions Including No Change

x_0 Bin	Party	Evidence Type	N	Mean PRE	$\bar{\Delta}_x$	Max Disp.	$\hat{\rho}$	Pos.	Neg.	No Chg.
2.5	1	Refutative	1123	0.821	17.256	97.5	0.177	0.536	0.094	0.370
2.5	1	Supportive	858	1.002	46.648	97.5	0.478	0.819	0.038	0.142
2.5	2	Refutative	878	0.883	18.432	97.5	0.189	0.556	0.076	0.368
2.5	2	Supportive	877	0.908	43.075	97.5	0.442	0.778	0.032	0.190
7.5	1	Refutative	435	8.605	11.393	92.5	0.123	0.545	0.421	0.034
7.5	1	Supportive	360	8.675	44.681	92.5	0.483	0.889	0.094	0.017
7.5	2	Refutative	371	8.531	14.825	92.5	0.160	0.623	0.340	0.038
7.5	2	Supportive	357	8.580	44.406	92.5	0.480	0.880	0.115	0.006
12.5	1	Refutative	354	13.138	10.282	87.5	0.118	0.542	0.441	0.017

Continued on next page

12.5	1	Supportive	288	13.184	42.576	87.5	0.487	0.913	0.087	0.000
12.5	2	Refutative	356	12.980	13.197	87.5	0.151	0.590	0.362	0.048
12.5	2	Supportive	301	13.123	43.973	87.5	0.503	0.900	0.093	0.007
17.5	1	Refutative	624	18.596	4.484	82.5	0.054	0.425	0.537	0.038
17.5	1	Supportive	501	18.615	41.425	82.5	0.502	0.870	0.110	0.020
17.5	2	Refutative	600	18.598	6.733	82.5	0.082	0.477	0.477	0.047
17.5	2	Supportive	533	18.627	34.859	82.5	0.423	0.827	0.144	0.028
22.5	1	Refutative	522	23.153	4.734	77.5	0.061	0.433	0.542	0.025
22.5	1	Supportive	434	23.145	37.422	77.5	0.483	0.859	0.136	0.005
22.5	2	Refutative	508	23.152	7.803	77.5	0.101	0.508	0.476	0.016
22.5	2	Supportive	453	23.099	36.790	77.5	0.475	0.863	0.128	0.009
27.5	1	Refutative	807	28.746	-1.414	72.5	-0.020	0.369	0.602	0.029
27.5	1	Supportive	683	28.704	31.177	72.5	0.430	0.814	0.165	0.020
27.5	2	Refutative	798	28.625	2.360	72.5	0.033	0.432	0.530	0.038
27.5	2	Supportive	727	28.579	31.182	72.5	0.430	0.807	0.176	0.017
32.5	1	Refutative	675	32.991	-0.797	67.5	-0.012	0.390	0.584	0.027
32.5	1	Supportive	555	32.996	27.386	67.5	0.406	0.766	0.218	0.016
32.5	2	Refutative	665	33.009	2.481	67.5	0.037	0.432	0.549	0.020
32.5	2	Supportive	568	32.942	26.470	67.5	0.392	0.755	0.232	0.012
37.5	1	Refutative	811	38.517	-7.202	62.5	-0.115	0.314	0.663	0.022
37.5	1	Supportive	782	38.586	23.836	62.5	0.381	0.756	0.229	0.015
37.5	2	Refutative	962	38.514	-3.952	62.5	-0.063	0.344	0.625	0.031
37.5	2	Supportive	887	38.643	23.418	62.5	0.375	0.770	0.210	0.020
42.5	1	Refutative	573	42.733	-8.864	57.5	-0.154	0.291	0.689	0.019
42.5	1	Supportive	502	42.733	21.657	57.5	0.377	0.765	0.227	0.008
42.5	2	Refutative	649	42.844	-6.974	57.5	-0.121	0.300	0.684	0.015
42.5	2	Supportive	573	42.864	19.133	57.5	0.333	0.733	0.237	0.030
47.5	1	Refutative	516	48.857	-12.717	52.5	-0.242	0.318	0.634	0.048

Continued on next page

47.5	1	Supportive	556	48.942	17.444	52.5	0.332	0.732	0.221	0.047
47.5	2	Refutative	575	48.920	-10.350	52.5	-0.197	0.310	0.631	0.059
47.5	2	Supportive	589	48.883	16.102	52.5	0.307	0.713	0.233	0.054
52.5	1	Refutative	1057	52.188	-11.610	52.5	-0.221	0.315	0.596	0.089
52.5	1	Supportive	1042	52.245	15.746	52.5	0.300	0.722	0.179	0.099
52.5	2	Refutative	1071	52.302	-10.380	52.5	-0.198	0.318	0.570	0.111
52.5	2	Supportive	1087	52.266	11.251	52.5	0.214	0.632	0.250	0.118
57.5	1	Refutative	593	58.642	-16.054	57.5	-0.279	0.339	0.639	0.022
57.5	1	Supportive	720	58.719	10.385	57.5	0.181	0.718	0.246	0.036
57.5	2	Refutative	619	58.685	-15.840	57.5	-0.275	0.326	0.654	0.019
57.5	2	Supportive	757	58.721	6.214	57.5	0.108	0.662	0.306	0.032
62.5	1	Refutative	781	62.863	-17.022	62.5	-0.272	0.335	0.636	0.028
62.5	1	Supportive	1072	62.812	8.447	62.5	0.135	0.719	0.261	0.020
62.5	2	Refutative	819	62.871	-18.087	62.5	-0.289	0.310	0.674	0.016
62.5	2	Supportive	898	62.811	5.687	62.5	0.091	0.643	0.317	0.040
67.5	1	Refutative	816	68.517	-21.091	67.5	-0.312	0.300	0.675	0.025
67.5	1	Supportive	1011	68.338	4.495	67.5	0.067	0.657	0.318	0.025
67.5	2	Refutative	861	68.407	-19.876	67.5	-0.294	0.300	0.671	0.029
67.5	2	Supportive	996	68.474	1.870	67.5	0.028	0.614	0.359	0.026
72.5	1	Refutative	984	72.841	-23.583	72.5	-0.325	0.276	0.702	0.021
72.5	1	Supportive	1270	72.810	1.294	72.5	0.018	0.600	0.376	0.024
72.5	2	Refutative	1038	72.739	-23.921	72.5	-0.330	0.262	0.724	0.014
72.5	2	Supportive	1260	72.787	-1.686	72.5	-0.023	0.546	0.428	0.026
77.5	1	Refutative	851	78.363	-26.536	77.5	-0.342	0.235	0.747	0.018
77.5	1	Supportive	1002	78.364	-1.691	77.5	-0.022	0.558	0.408	0.034
77.5	2	Refutative	837	78.321	-28.174	77.5	-0.364	0.211	0.772	0.017
77.5	2	Supportive	906	78.312	-2.930	77.5	-0.038	0.517	0.450	0.033
82.5	1	Refutative	916	82.713	-30.226	82.5	-0.366	0.221	0.758	0.022

Continued on next page

82.5	1	Supportive	1077	82.627	-3.917	82.5	-0.047	0.513	0.460	0.027
82.5	2	Refutative	906	82.692	-32.304	82.5	-0.392	0.174	0.807	0.019
82.5	2	Supportive	1072	82.669	-7.464	82.5	-0.090	0.433	0.533	0.035
87.5	1	Refutative	701	88.248	-35.755	87.5	-0.409	0.175	0.816	0.009
87.5	1	Supportive	757	88.273	-6.317	87.5	-0.072	0.474	0.490	0.036
87.5	2	Refutative	653	88.204	-34.907	87.5	-0.399	0.139	0.847	0.014
87.5	2	Supportive	678	88.176	-10.383	87.5	-0.119	0.375	0.591	0.034
92.5	1	Refutative	718	92.756	-40.015	92.5	-0.433	0.146	0.834	0.019
92.5	1	Supportive	711	92.578	-9.609	92.5	-0.104	0.421	0.532	0.048
92.5	2	Refutative	670	92.770	-42.031	92.5	-0.454	0.096	0.887	0.018
92.5	2	Supportive	670	92.628	-12.628	92.5	-0.137	0.331	0.643	0.025
97.5	1	Refutative	2415	99.554	-41.921	97.5	-0.430	0.029	0.754	0.216
97.5	1	Supportive	2091	99.543	-12.858	97.5	-0.132	0.069	0.476	0.454
97.5	2	Refutative	1878	99.566	-44.780	97.5	-0.459	0.021	0.778	0.201
97.5	2	Supportive	1525	99.479	-15.691	97.5	-0.161	0.050	0.555	0.395

Table 32: Descriptive Statistics by Bin, Party, and Evidence Type

T^1 Bin	Party	Evidence Type (E_S, E_R)	N	Mean T^1	$\bar{\Delta}T$	SD ΔT	Mean $\hat{\rho}$	SD $\hat{\rho}$	Prop. $+\Delta$	Prop. $-\Delta$	Prop. No Δ
$[0,25]$	1	Ref.	3058	10.79	10.87	27.04	0.117	0.302	0.498	0.348	0.155
$[0,25]$	1	Sup.	2441	11.12	43.17	33.80	0.487	0.376	0.858	0.084	0.057
$[0,25]$	2	Ref.	2713	11.60	12.67	27.10	0.139	0.306	0.543	0.313	0.144
$[0,25]$	2	Sup.	2521	11.19	40.50	34.07	0.457	0.380	0.833	0.092	0.075
$(25,50]$	1	Ref.	3382	37.37	-5.67	25.83	-0.100	0.425	0.339	0.633	0.028
$(25,50]$	1	Sup.	3078	37.93	24.60	26.87	0.393	0.433	0.768	0.211	0.021
$(25,50]$	2	Ref.	3649	37.76	-2.94	26.03	-0.057	0.429	0.366	0.602	0.032
$(25,50]$	2	Sup.	3344	38.01	23.60	26.72	0.376	0.429	0.759	0.215	0.026
$(50,75]$	1	Ref.	4231	63.02	-17.84	28.40	-0.279	0.455	0.310	0.649	0.040
$(50,75]$	1	Sup.	5115	63.66	7.65	22.32	0.132	0.366	0.678	0.282	0.040
$(50,75]$	2	Ref.	4408	63.12	-17.62	27.93	-0.274	0.444	0.301	0.657	0.042
$(50,75]$	2	Sup.	4998	63.54	4.36	23.46	0.079	0.382	0.613	0.337	0.049
$(75,100]$	1	Ref.	5601	91.29	-36.65	35.69	-0.399	0.386	0.125	0.772	0.103
$(75,100]$	1	Sup.	5638	90.16	-7.88	22.73	-0.083	0.253	0.340	0.470	0.190
$(75,100]$	2	Ref.	4944	90.46	-38.01	34.63	-0.417	0.376	0.107	0.806	0.087
$(75,100]$	2	Sup.	4851	89.28	-10.32	23.49	-0.112	0.262	0.306	0.548	0.146

Statistic	**Mean Excess Change**	**t-Statistic**	**p-Value**
Estimate	-2.816	-22.118	< 0.00001

Table 33: Results of Lord–Novick Excess Change Test.

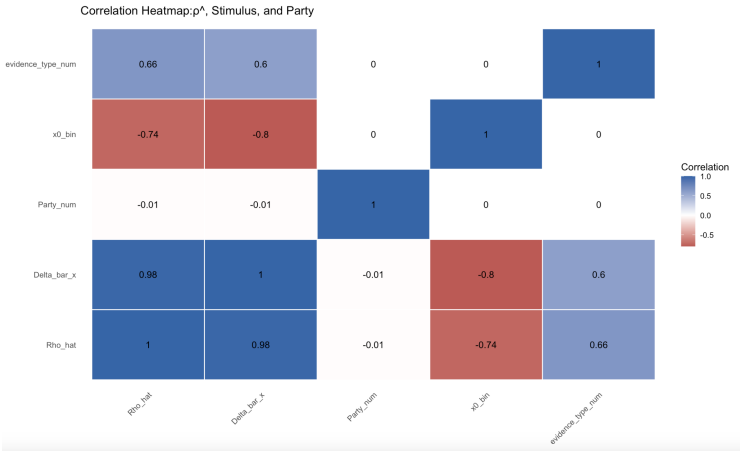


Figure 24: Section 2, Correlation Matrix

Table 34: Binomial Test Results for Directional Change at Scale Extremes

Quartile	Trials (n)	Successes (k)	$\hat{p}$	95% CI (Clopper–Pearson)
Q1 (Positive Shift)	16,211	10,379	0.640	[0.634, 1.000]
Q4 (Negative Shift)	15,389	10,287	0.668	[0.662, 1.000]

Note. Binomial tests were conducted with the null hypothesis $H_0 : p = 0.5$. For both tests, the alternative hypothesis was that the true probability of directional change exceeds chance ($p > 0.5$). Both tests yielded $p < 2.2 \times 10^{-16}$, indicating highly significant deviations from chance.

Table 35: ANOVA Results for Absolute Belief Change by Initial Position Quartile

Source	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
PRE Quartile	3	473,025	157,675	260.9	$< 2 \times 10^{-16}$***
Residuals	63,968	38,659,147	604		

(2) Original Research Siegrist & Bearth, 2021

Psychological variables—particularly trust and worldviews—significantly shape individuals' risk perceptions and acceptance of protective measures. The authors demonstrate these effects using both cross-sectional and longitudinal designs.

While the number of infections declined between survey waves 1 and 2, longitudinal analyses support the conclusion that changes in perceived health risk, social trust, and the perception of other risks are associated with changes in the acceptance of policy measures. A limitation is that the data were collected solely in Switzerland. Pandemic phase, governmental measures, public compliance, sociocultural context, and economic conditions may influence not only overall acceptance but also the salience of specific explanatory factors in shaping risk perceptions and behavioral responses.

Results of a longitudinal linear regression analysis with acceptance of measures in survey wave 2 as the dependent variable

	Unstandardized B	SE	Beta	*t*
Constant	3.54	0.35		10.15**
Sex[†]	−0.11	0.06	−0.03	−1.93
Age	−0.001	0.002	−0.01	−0.62
Risk group[‡]	0.04	0.07	0.01	0.53
Acceptance of measures (T1)	0.56	0.03	0.48	20.00**
Individualism (T2)	−0.28	0.03	−0.22	−9.57**
Hierarchy (T2)	−0.04	0.03	−0.03	−1.61
General interpersonal trust (T2)	−0.08	0.03	−0.06	−3.25*
Social trust (T1)	0.06	0.03	0.06	2.28
Perceived health risks (T1)	0.10	0.03	0.09	3.91**
Perceived costs of COVID-19 measures (T1)	−0.11	0.03	−0.11	−4.06**
Social trust, change score (T2 – T1)	0.10	0.04	0.06	2.97*
Perceived health risks, change score (T2 – T1)*	0.23	0.03	0.14	7.54**
Perceived costs of COVID-19 measures, change score (T2 – T1)	−0.16	0.03	−0.14	−6.00**

$R^2 = 0.64$, T1 = variable wave 1, and T2 = variable wave 2. *$P < 0.01$, **$P < 0.001$.
[†]Sex: male coded as 0; female coded as 1.
[‡]Belonging to objective risk group: no coded as 0; yes coded as 1.

Figure 25: Table 3 from the original study presents all measured variables; that which is marked with a red asterisk was included in the secondary analysis.

Perceived Health Risks Attributed to SARS-CoV-2

Participants' perceptions of health-related risks associated with SARS-CoV-2 were assessed using five items. Respondents rated their level of concern on a 7-point Likert scale (1 = no fear, 7 = very high fear), with only the endpoints labeled descriptively. Items addressed fears of personal infection, infection of family or acquaintances, fatalities in one's social environment, widespread fatalities in Switzerland, and overburdening of the healthcare system. The scale demonstrated high internal consistency, with Cronbach's α values of 0.87 (wave 1) and 0.88 (wave 2).

(3) Original Research Anglin, 2019 (Study 2 of 4)

The original research examined how individuals' beliefs about the relationship between religiosity and life outcomes responded to empirical evidence, focusing on the roles of prior belief, perceived evidence quality, and belief recall. Participants (N = 427) completed belief ratings at baseline (T1), after exposure to research findings (T2), and recalled their T1 beliefs during the second session. Belief ratings were measured

on a 9-point scale (1 = strongly negative, 9 = strongly positive outcomes). Mean belief scores slightly declined from T1 (M = 5.97) to T2 (M = 5.76), with an average belief change of 0.73. Religiosity (M = 3.21) and religious group identification (M = 4.09) were also assessed.

Participants reviewed ten research summaries (two each for psychological well-being, vocational outcomes, and cognitive abilities; three for social support; one for personality), of which eight showed a directional effect (favoring either religious or non-religious individuals) and two showed no difference. A multiple-choice attention check followed each summary. No participants were excluded for failing attention checks, given the high overall accuracy (93%).

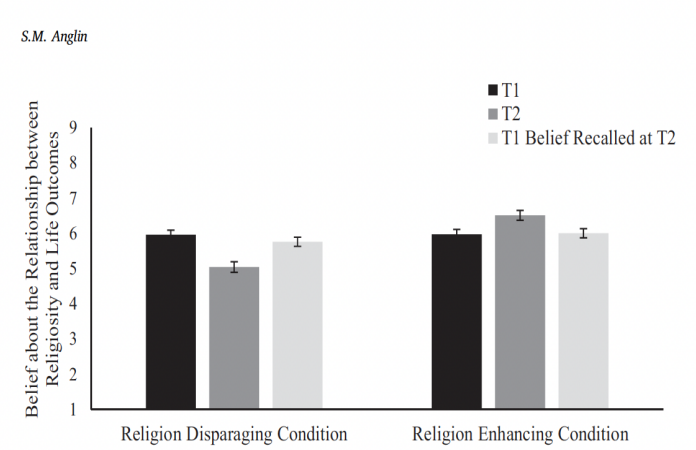


Figure 26: Figure 2 from the original study. Variables $T1$ and $T2$ are used in the secondary analysis.

The original findings suggest that while biased evaluation of evidence partly contributed to belief maintenance, belief change also occurred and was modestly associated with perceived evidence quality. Importantly, belief change was not significantly moderated by whether the evidence was congruent or incongruent with participants' initial beliefs, indicating that evaluation of evidence influenced belief revision independently of prior convictions.