

Decisions under Narrative Complexity*

Mustafa Kaba¹, Simay Küçükkolbaşı², and Cavit Görkem Destan³

¹Koç University, İstanbul, Turkey

²Vienna University of Economics & Business

³TED University, Ankara, Turkey

May 13, 2026

Abstract

How does the structural architecture of a causal narrative shape whether individuals correctly comprehend and act upon it? We design a pre-registered online experiment that decomposes narrative complexity into two dimensions: *state complexity*, the number of variables in a linear causal chain, and *structure complexity*, the presence of a negative feedback loop introducing mutual causation. Structure complexity imposes markedly higher cognitive costs than state complexity despite involving fewer variables. A four-variable chain generates no statistically significant effect relative to baseline, while a five-variable chain raises comprehension errors by 23 percent and decision errors by 17 percent. The feedback loop raises comprehension errors by 41 percent and decision errors by 21 percent—effects substantially larger despite the feedback loop containing fewer variables than the five-variable chain. These results reveal a fundamental asymmetry between the *simulation costs* of processing interdependent systems and the *storage costs* of tracking longer causal chains. These findings identify a fundamental tension in policy communication: feedback loops are endemic to accurate representations of general equilibrium effects, making factually complete policy arguments harder to process than simplified linear narratives.

JEL Codes: C91, D83, D91

Keywords: narratives; causal structures; complexity; cognition

*The study was pre-registered on the AEA RCT Registry under *AEARCTR-0017232*. Mustafa Kaba gratefully acknowledges funding from the *EU Horizon's Marie Skłodowska-Curie Actions (MSCA) Postdoctoral Fellowship Program* and the *International Fellowship for Early Stage Researchers Program of TÜBİTAK* (Scientific and Technological Research Council of Turkey). The IRB approval was received from the Social Science Research Ethics Committee of Koç University.

1 Introduction

“Human rational behavior is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor.”

Herbert Simon, 1990

Modern democracies ask citizens to evaluate complex causal claims. Debates over immigration, trade, climate policy, and monetary intervention hinge on causal chains—how a policy intervention A affects variable B , which in turn shapes outcome C —with the full argument sometimes involving feedback loops, multiple intermediate variables, and equilibrium effects that can reverse the apparent direction of a simple causal link. These arguments reach citizens primarily as *causal narratives*: verbal representations of cause-and-effect chains communicated through news coverage, political speeches, and social media (Shiller, 2020; Andre et al., 2022; Blesse et al., 2025). In Simon’s terms, the causal narrative is the task environment; the human cognitive apparatus is the computational blade. Whether citizens correctly comprehend a policy argument—and act on it—depends on both.

Traditional research on belief updating in economics typically assumes that individuals know the structure of their environment and asks whether they update by the correct magnitude (Benjamin, 2019). The more fundamental question—whether people correctly represent the causal structure at all—however has received far less attention, especially in complex environments (Ambuehl et al., 2026). Economists have begun to document the behavioral consequences of such misrepresentation: Dal Bó et al. (2018) show experimentally that voters systematically underestimate general equilibrium effects of policy interventions, generating widespread support for welfare-reducing policies, and Nunnari et al. (2024) demonstrate that lower cognitive ability amplifies this failure. These findings establish that miscomprehension of causal structure—not only partisan preferences—is a first-order driver of policy opinion. Yet neither the precise source of cognitive difficulty nor the structural features of causal narratives that generate it have been characterized.¹

We implement a pre-registered, mixed-factorial online experiment ($N = 1,800$, representative US sample) to causally identify the effects of two theoretically distinct dimensions of narrative complexity on comprehension and decision quality. Our central design innovation

¹For a comprehensive review of the economics and cognitive science of causal mental models, see Ambuehl et al. (2026). For the role of competing causal narratives in generating policy disagreement and populism, see Eliaz and Spiegler (2020) and Levy et al. (2022).

is a structural decomposition of complexity grounded in cognitive load theory and the cognitive science of causal reasoning. The first dimension—*state complexity*—varies the *length* of the causal chain, progressively adding variables to a baseline linear narrative. The second dimension—*structure complexity*—varies the *nature* of causal interactions by introducing a negative feedback loop into an otherwise short chain. This decomposition allows us to separately identify the *storage costs* of tracking longer causal chains from the *simulation costs* of mentally resolving a system with mutually determined variables.

In the experiment, we randomly assign participants to one of four complexity conditions: a baseline three-variable chain ($a \rightarrow b \rightarrow c$), corresponding to the archetypal causal structure studied extensively in the cognitive science of causal reasoning (Rottman and Hastie, 2014; Ambuehl and Thyssen, 2025); a four-variable extension (State v1: $a \rightarrow b \rightarrow c \rightarrow d$); a five-variable branching chain (State v2: $a \rightarrow b \rightarrow c \rightarrow d$, $c \rightarrow e$); or a three-variable chain with a negative feedback loop (Structure: $a \rightarrow b \leftrightarrow c$). Each participant then completes incentivized tasks across three domains—fictional (abstract labels designed to eliminate prior beliefs), production (industrial context), and immigration (politically charged, with a randomized Democrat- or Republican-leaning frame)—spanning a deliberate spectrum from clean experimental control to real-world policy relevance. Subjects study each narrative for two minutes, take notes in a constrained text field following Fréchette et al. (2024), and then complete: (i) an incentivized Decision Task in which taking the action is always optimal, and (ii) a Rulebook questionnaire probing comprehension of all potential pairwise causal links.

We embed each causal structure in a concrete empirical cover story that translates the abstract graph into a coherent verbal narrative with domain-specific variables. In the production domain, for instance, the baseline chain ($a \rightarrow b \rightarrow c$) reads: “An increase in the *availability of cheaper energy* (a) leads to a higher rate of *automated production* (b); in turn, a higher *automated production* (b) rate leads to a decrease in the *defect rate* (c).” Structurally identical narratives instantiate the same underlying graph in the fictional and immigration domains, varying only the variable labels. This design reflects the medium through which people most naturally transmit and receive causal knowledge. People primarily exchange causal information through verbal explanations rather than formal diagrams (Sloman et al., 2009; Lombrozo and Vasilyeva, 2017; Meder and Mayrhofer, 2017), making our textual format both ecologically valid and the relevant benchmark for understanding how narrative complexity shapes comprehension and decision-making.

The two complexity manipulations target distinct cognitive resources. State complexity features *low element interactivity*: causal links admit sequential processing, so the cognitive

burden falls primarily on storage, i.e., working memory.² Structure complexity features *high element interactivity*: the variables in a feedback loop are co-determined, so agents cannot process them sequentially but must mentally simulate a dynamic system to compute net effects, i.e., agents must *equilibrate* to find a fixed point (Ba et al., 2024; Ba et al., 2025). Cognitive science establishes that people strongly prefer linear causal reasoning and systematically fail to “close the loop” when confronting feedback structures, treating $b \leftrightarrow c$ as if it were $b \rightarrow c$ (Kastens and Shipley, 2023). The feedback loop thus imposes not just a storage cost but a simulation cost—a qualitatively distinct and more demanding cognitive operation.

Three main findings emerge. First, both dimensions of complexity raise error rates relative to baseline. The four-variable chain (State v1) generates no statistically significant effects, consistent with this length remaining within typical working memory capacity. The five-variable chain (State v2) raises comprehension errors by 4.1 percentage points and decision errors by 3.6 percentage points—23 and 17 percent above baseline means of 18 and 21 percent, respectively—revealing a threshold phenomenon consistent with working memory limits (Halford et al., 2005). Second, structure complexity dominates state complexity as a source of cognitive cost despite the feedback loop narrative containing *fewer* variables than State v2. The feedback loop raises comprehension errors by 7.5 percentage points and decision errors by 4.4 percentage points—41 and 21 percent above baseline—revealing a fundamental asymmetry: the simulation costs of interdependent causal structures substantially exceed the storage costs of longer causal chains.

Third, self-reported cognitive uncertainty and perceived difficulty increase monotonically with complexity, providing validation for the experimental manipulations. A particularly striking additional finding concerns domain heterogeneity: in the immigration domain, complexity effects on error rates essentially vanish as prior political beliefs displace causal reasoning, regardless of narrative structure. We find that participants who identify themselves as Republicans exhibit higher error rates in almost all domains and complexity levels compared to their fellow Democrats.

We decompose the main results into two distinct cognitive failure modes. *Representational failure*—the inability to correctly encode which variables and causal links exist—stems from memory overload as chain length exceeds working memory capacity. *Computational fail-*

²Psychological research identifies a natural working memory capacity limit of approximately *four* variables (Luck and Vogel, 1997; Halford et al., 2005; Bays et al., 2011; Oberauer et al., 2016; Bohren et al., 2024; Loewenstein and Wojtowicz, 2025); our design probes this threshold directly with chains of three, four, and five variables.

ure—the inability to correctly compute net causal effects from an accurate representation—stems from the simulation cost of resolving a feedback loop. We isolate these channels using three complementary strategies: (i) treatment effects on errors for a computationally trivial Rulebook question (a baseline link present in all treatment arms), which state complexity elevates but structure complexity does not; (ii) within-treatment variation in error rates across domains, where fictional labels—which provide no real-world memory anchor—generate the highest errors in every treatment arm; and (iii) conditioning on representational accuracy, which substantially shrinks the State v2 coefficient but leaves the Structure coefficient largely unchanged. This convergent evidence confirms that state and structure complexity operate through qualitatively distinct cognitive channels.

This study speaks to several strands of the literature. A growing literature studies how individuals form and act upon mental models of causal structures.³ The papers closest to ours are [Fréchette et al. \(2024\)](#), [Kendall and Oprea \(2024\)](#), [Ambuehl and Thyssen \(2025\)](#), and [Charles and Kendall \(2025\)](#), which investigate mental models built from abstract observational data or choices among causal diagrams. Our paper differs along two key dimensions. First, we hold the informational content of the narrative fixed and vary its structural complexity, providing clean identification of purely structural effects on comprehension and decision quality. Second, we work exclusively with textual narratives—the primary medium through which citizens receive causal information. This choice reflects the finding of [Nam et al. \(2023\)](#) that participants learn better from explanations than from observations. By delivering causal structure in explicit textual form rather than requiring subjects to infer it from data, we create conditions more favorable to comprehension. In this sense, our study provides a conservative lower bound on the cognitive costs of complexity: unlike the closest related papers, which require subjects to infer causal structure from abstract observational data ([Fréchette et al., 2024](#); [Charles and Kendall, 2025](#)), we deliver explicit causal information in standard textual form—and still document substantial comprehension and decision failures.

A further novelty is that our structure complexity treatment introduces a feedback loop—a feature that standard causal DAG frameworks accommodate only by unfolding the system over time ([Ambuehl et al., 2026](#))—and we show that this architectural feature imposes qualitatively larger cognitive costs than a longer linear chain. We also contribute to the literature on narratives and economic beliefs ([Shiller, 2020](#); [Andre et al., 2022](#); [Kendall and Charles, 2022](#); [Andre et al., 2024](#); [Blesse et al., 2025](#)) by shifting focus from narrative *content*

³See [Ambuehl et al. \(2026\)](#) for a recent review.

to narrative *structure*: the causal architecture of a narrative—not only what it says but how it organizes cause and effect—is a first-order determinant of comprehension and decision quality.

A recent program of research examines how task complexity affects economic decisions (Kalaycı and Serra-Garcia, 2016; Oprea, 2020; Banovetz and Oprea, 2023; Arrieta and Nielsen, 2024). Oprea (2024) shows that violations of expected utility theory emerge from complexity manipulations involving no risk. Gabaix and Graeber (2024) provide a general theoretical framework, and Agranov et al. (2025) document substantial individual heterogeneity in subjective complexity perceptions. We extend this program to the comprehension of policy-relevant causal narratives in verbal form—a domain of high external validity—and provide the first two-dimensional structural decomposition of narrative complexity that maps onto distinct cognitive mechanisms. Where existing work typically treats complexity as a single aggregate cost, we show that cognitive cost has at least two separable dimensions—storage and simulation costs—that predict qualitatively different failure modes and call for different corrective interventions. A memory aid might alleviate the storage costs of a longer chain; a structural visualization may be necessary to reduce the simulation costs of a feedback loop.

Finally, Dal Bó et al. (2018) show that voters who fail to trace the general equilibrium effects of policy interventions demand welfare-reducing policies, and Nunnari et al. (2024) demonstrate that lower cognitive ability amplifies this failure. Our experiment supplies the missing structural micro foundation: the feedback loop is precisely the architectural feature of causal narratives that imposes the highest cognitive cost, and subjects who fail to process it might make decision errors even when they correctly encode the underlying variables. This provides a direct experimental link between structural narrative complexity and the failure to appreciate general equilibrium consequences—the cognitive root of demand for bad policy. More broadly, because feedback loops are endemic to accurate representations of general equilibrium effects, factually complete policy communication is inevitably harder to process than simplified linear narratives that strip out recursive consequences (MacKay, 2003; Clippel et al., 2024; Kendall and Oprea, 2024). Our results thus reveal a fundamental tension between accuracy and accessibility in policy communication that is structural rather than incidental.

The remainder of the paper proceeds as follows. Section 2 describes the experimental design. Section 3 describes the data and variables. Section 4 presents the main results, mechanisms, and domain-level heterogeneity. Section 5 concludes.

2 Experimental Design

To isolate the effects of narrative complexity, we implement a pre-registered, mixed factorial experiment (3×3 design) involving a representative sample of the US population ($N = 1800$). More specifically, the design varies the complexity of causal narratives (*2 complexity variations + 1 control arm*) presented to subjects while holding the optimal decision constant across three different domains (*fictional, production, political:immigration*) and with two different political slants (*Democrat vs. Republican*).

Between-Subjects Factor: Narrative Complexity. The primary experimental manipulation involves Narrative Complexity, which is varied between subjects. Each participant is randomly assigned to one of three complexity treatment arms:

- *Baseline:* Assigned to one-third of the sample.
- *State Complexity:* Assigned to one-third of the sample, further divided equally into variations v1 and v2.
- *Structure Complexity:* Assigned to the final one-third of the sample.

Within-Subjects Factor: Topic Domain. To assess the robustness of complexity effects across different contexts, we include Topic Domain as a within-subjects factor. Each participant completes the task for all three of the following topics, presented in a randomized order:

- *Fictional:* Scenarios involving non-real-world entities like “aetheric energy flux” and “chrono-synaptic activation” designed to ensure respondents have no existing priors or biases.
- *Production:* Industrial narratives focused on automated production and defect rates.
- *Immigration:* Policy narratives featuring either Democrat-slanted or Republican-slanted frames.

To demonstrate the conceptual robustness of our complexity manipulations, we implement the experiment across three distinct domains chosen for their varying degrees of real-world relevance and potential for prior beliefs. At one end of the spectrum, the fictional

setting serves as an artificial, “clean” environment specifically designed to ensure respondents have no existing priors or biases. Moving toward greater realism, the production setting introduces a neutral, technical context where, although more grounded in real-life scenarios, individuals are unlikely to have strong priors. Finally, the immigration topic deliberately trades off these clean experimental conditions for high relevance to public policy and contemporary social discourse. This spectrum—spanning from a neutral “no-priors” baseline to a highly polarized real-world context—allows us to investigate whether the behavioral consequences of narrative complexity remain consistent regardless of the underlying topic’s familiarity or ideological weight.

In the Immigration domain, we manipulate the narrative slant to reflect common political framing. These baseline examples illustrate how the same underlying structure ($a \rightarrow b \rightarrow c$) is adapted:

- *Republican Leaning*: “A recent analysis shows that an increase in immigration restrictions leads to a rise in social integration of immigrants. In turn, a rise in social integration of immigrants leads to a decrease in the crime rate of foreigners. No other effects are observed.”
- *Democrat Leaning*: “A recent analysis shows that an increase in open immigration policies leads to a rise in cultural diversity. In turn, a rise in cultural diversity leads to a decrease in social isolation. No other effects are observed.”

Each participant reads exactly one narrative in each domain (Fictional, Production, Immigration). Crucially, each narrative is presented with a different complexity level, meaning a participant reads a total of three unique complexity-domain pairs. This ensures that every participant experiences all complexity and domain variations exactly once during the session. All resulting treatment effects are compared against the baseline narrative performance.

Theoretical Motivation for Complexity Manipulations. We manipulate complexity along two primary dimensions over the baseline structure:

0. **Baseline** ($s = 3$): The baseline condition serves as a benchmark for simple, linear causal narrative processing. It consists of a straightforward linear chain with three variables ($a \rightarrow b \rightarrow c$), where each link can be processed sequentially.
1. **State Complexity**: First manipulation features *low element interactivity*. A subject can process the $a \rightarrow b$ link, consolidate that information, and then proceed to the

$b \rightarrow c$ link. The causal relationships can be understood sequentially without holding all elements in working memory simultaneously. Psychological literature suggests a processing threshold around four variables (Halford et al., 2005):

- *State v1* ($s = 4$): The baseline chain is extended by one variable ($a \rightarrow b \rightarrow c \rightarrow d$), adding one new variable and one new causal link.
- *State v2* ($s = 5$): The chain is extended by two variables, creating a branching effect ($a \rightarrow b \rightarrow c \rightarrow d$ and $c \rightarrow e$).

2. **Structure Complexity:** Second manipulation features *high element interactivity*, taxing the ability of working memory to perform mental simulations of a dynamic system. To understand the net effect of a change in variable b , the subject must simultaneously consider its direct effect on c and the subsequent recursive effect of c back on b . Because these elements are co-determined, they cannot be processed in isolation and must be considered jointly to compute the net effect:

- *Structure Treatment:* A negative feedback loop is introduced into the baseline structure ($a \rightarrow b \leftrightarrow c$), creating a cyclical causality that requires significantly deeper cognitive processing to unravel compared to linear chains.

These complexity variations are summarized in Table 1.

Table 1: Treatment Arms: Complexity Manipulations

Arm	Type	Description	Structure
Baseline	—	Simple causal chain	$a \rightarrow b \rightarrow c$
State v1	State complexity	Long chain	$a \rightarrow b \rightarrow c \rightarrow d$
State v2	State complexity	Longer chain	$a \rightarrow b \rightarrow c \rightarrow d, c \rightarrow e$
Structure	Structure complexity	Feedback loop	$a \rightarrow b \leftrightarrow c$

Notes: State complexity extends the length of the causal chain, imposing storage costs. Structure complexity introduces a feedback loop, imposing simulation costs.

The experimental flow for each topic consists of:

1. **Briefing:** Subjects study a short narrative for 2 minutes and are prompted to take notes (70-character limit) as in Fr chet te et al. (2024).

2. **Decision Task (incentivized):** Starting with a \$10 bonus, subjects must decide whether to pay \$5 to take an action (e.g., increase Social Integration of Immigrants) to achieve a goal (e.g., decrease Crime Rate of Foreigners). Taking the action is the optimal choice in all treatments.
3. **Rulebook Questions (incentivized):** A series of multiple-choice questions tests the comprehension of all potential causal links. In State treatments, the first six questions are fixed (regarding a , b , c) to ensure comparability.
4. **Self-Report:** Participants report perceived difficulty, confidence (cognitive uncertainty), and effort on 1–10 scales.

3 Data

In this section, we first describe the survey logistics, and second, explain in detail the main variables used in the analyses.

3.1 Survey Logistics

The experiment was administered online via Qualtrics to a sample of $N = 1,800$ participants recruited from Prolific. To ensure external validity, the sample is representative of the U.S. population in terms of sex, age, and political affiliation, according to U.S. Census data. The experiment was conducted in January–February 2026.

3.2 Outcome Variables

We collect two primary outcome measures that capture *experienced complexity*—the actual cognitive difficulty imposed on subjects by the narrative structure.

Decision Task error rate. The Decision Task asks subjects to decide whether to pay \$5 to increase variable b (the action variable), with the goal of decreasing variable c (the target). Crucially, the causal diagrams in all treatments are constructed such that taking the action is always the optimal decision—that is, the net causal effect of b on c is strictly negative in every treatment arm, including the structure treatment with the feedback loop. An error is recorded when the subject declines to take the action. We use the binary error indicator

as our primary outcome: 1 if the subject made the suboptimal choice (no action), 0 if they made the optimal choice (pay \$5 and act).

Rulebook error rate. The Rulebook task exhausts all potential pairwise causal relationships between the variables in the narrative, including relationships that do not exist. Subjects must correctly identify present causal links (with their correct sign) and absent causal links. The first six questions are held constant across all treatment arms and concern the relationships among the core variables a , b , and c . In State treatments, additional questions are added about newly introduced variables (d and e). Our main Rulebook outcome is the error rate on the question about the $\rightarrow c$ link (RB Q4 in our coding), which probes the direct causal link between the action variable b and the target variable c —essentially asking the same inferential question as the Decision Task but in a more structured format. We also analyze average error rates across all six fixed questions as a broader measure of comprehension.

Notes. During the 2-minute briefing period, subjects are invited to take free-form notes in a text field with a 70-character limit, following the protocol of [Fr chet te et al. \(2024\)](#). We record whether a subject took any notes (a binary indicator) and analyze the content of notes to explore reasoning processes and the cognitive strategies subjects employ under different complexity conditions.

3.3 Secondary Dependent Variables

Following each narrative task, subjects report three self-assessed measures on 1–5 scales:

Perceived Complexity (PC): “How difficult did you find the task?” (1 = very easy, 5 = very difficult). This captures subjects’ *perceived* difficulty, which may diverge from their *experienced* difficulty as measured by error rates.

Cognitive Uncertainty (CU): “How confident are you in your answers?” (1 = not confident at all, 5 = fully confident). We reverse the scale so that higher values indicate greater uncertainty, consistent with the interpretation of CU as a measure of subjective complexity.

Effort: “How much effort did you put into the task?” (1 = no effort, 5 = maximum effort).

We also record page-submission response times for the briefing page as an additional

measure of the time subjects allocate to processing the narrative. Following [Agranov et al. \(2025\)](#), we construct an awareness-of-complexity measure as the difference between standardized perceived complexity (PC) and standardized actual error rate, capturing the extent to which subjects recognize their own comprehension difficulties.

3.4 Control Variables

We measure fluid intelligence using five items from Raven’s Progressive Matrices, a standard non-verbal test of abstract reasoning. We compute a total IQ score (0–5) as the number of correct answers. IQ is expected to moderate the effects of complexity, particularly structure complexity, which requires higher-order simulation operations. Additionally, we collect standard demographic controls including age, gender, income category, education level, and political affiliation (party identification). These controls are used in alternate specifications to examine the robustness of main results and to explore heterogeneity.

We record subjects’ primary information sources (social media, printed newspaper, television, etc.) to capture variation in media exposure that may interact with narrative processing. Finally, we elicit topic-specific prior knowledge for the production and immigration domains. This serves as a proxy for the priors that subjects bring to the experimental session from real-world experience, allowing us to test whether greater familiarity with a topic mitigates or amplifies complexity effects.

4 Results

The analysis uses OLS regressions to estimate the impact of complexity treatments on error rates and self-reported metrics. To estimate the impact of narrative complexity on comprehension and decision quality, we employ the following econometric specification:

$$Y_{id} = \alpha + \beta_1 \text{ComplexityTreatment}_{id} + \delta_d + \psi_i + \gamma \mathbf{X}_i + \epsilon_{id}, \quad (1)$$

where Y_{id} is the primary dependent variable for respondent i in domain d , representing either the *Rulebook Question error rate* or the *Decision Task error rate*. $\text{ComplexityTreatment}_{id}$ is a treatment indicator for the assigned complexity level—*State-1*, *State-2*, *Structure*—with the *Baseline* narrative serving as the omitted category. δ_d represents domain fixed effects (*Fictional*, *Production*, *Immigration:Dem*, *Immigration:Rep*) to account for topic-specific variation. ψ_i represents respondent fixed effects. \mathbf{X}_i is a vector of individual-level

demographic controls, including *gender*, *age*, *income*, *education*, *political affiliation*, and *IQ score*. ϵ_{id} is the error term, with standard errors clustered at the respondent level to account for within-subject correlation across the three tasks. When we include individual-level controls, we drop respondent fixed effects ψ_i to avoid multicollinearity.

4.1 Primary Outcomes: Error Rates

Table 2 and Figure 1 report the main results. State-1 ($s = 4$) has a very limited and non-robust effect on error rates compared to the baseline narrative, consistent with working memory research suggesting that four-variable chains remain within typical cognitive capacity (Halford et al., 2005). State-2 ($s = 5$), on the other hand, shows a large and statistically significant effect on both Rulebook (0.041^{***}) and Decision Task errors (0.036^{**}). Given that mean error rates in the baseline arm are approximately 18 percent for the Rulebook and 21 percent for the Decision Task, these effect sizes represent a 23 percent and 17 percent increase over baseline means, respectively. This non-linearity—a null effect at four variables followed by a significant effect at five—provides clean evidence of a threshold phenomenon in the representational channel of cognitive difficulty.

The negative feedback loop (Structure) imposes the highest cognitive cost, leading to a 7.5 percentage point increase in Rulebook errors and a 4.4 percentage point increase in Decision Task errors. These effect sizes imply a 41 percent and 21 percent increase over baseline mean error rates. Crucially, these effects are substantially larger than the State-2 effects despite the Structure narrative containing *fewer* variables ($n = 3$ vs $n = 5$). This asymmetry is consistent with the theoretical prediction that structural endogeneity imposes simulation costs that are qualitatively more burdensome than the storage costs of tracking longer causal chains. Both the Rulebook and Decision Task results are robust to the inclusion of respondent fixed effects and demographic controls.

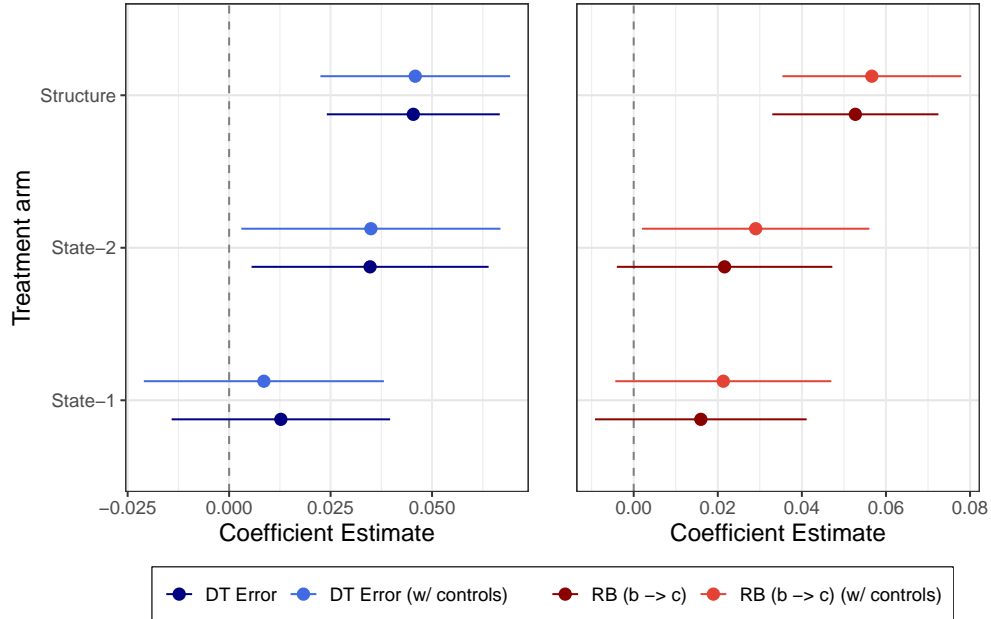
Domain variation. A key finding from the domain-level analysis is that complexity effects on error rates are largely driven by the fictional and production domains. In the immigration domain, complexity treatment effects on error rates are substantially smaller and often statistically indistinguishable from zero. This pattern suggests that prior political beliefs—which are particularly strong and salient in the immigration setting—dominate the processing of structural information when topics are politically charged. Given this finding, subsequent analyses of the cognitive mechanisms focus on the fictional and production domains unless otherwise noted.

Table 2: OLS Results for Rulebook and Decision Task Error Rates

	Error Rates			
	Decision Task		RB (b -> c)	
	(1)	(2)	(3)	(4)
State-1	0.013 (0.014)	0.009 (0.015)	0.016 (0.013)	0.021 (0.013)
State-2	0.035** (0.015)	0.035** (0.016)	0.022* (0.013)	0.029** (0.014)
Structure	0.045*** (0.011)	0.046*** (0.012)	0.053*** (0.010)	0.057*** (0.011)
Num.Obs.	1801	1511	1801	1511
Mean (Base. nar.)	0.21	0.21	0.11	0.11
Domain FE	Yes	Yes	Yes	Yes
Respondent FE	Yes	No	Yes	No
Demographic Controls	No	Yes	No	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Error Rates in Rulebook and Decision Task by Complexity Treatment



Notes: Coefficient plot from OLS regressions of error rates on complexity treatment indicators, with Baseline as omitted category. Confidence intervals at 95%. Standard errors clustered by respondent.

4.2 Secondary Outcomes: Self-Reported Measures

The experimental manipulation successfully shifted participants’ internal states along all measured dimensions except effort.

Cognitive Uncertainty (CU) and Perceived Complexity (PC). All three complexity treatments significantly increased both reported cognitive uncertainty and perceived complexity relative to baseline. All treatments—State-1, State-2, and Structure—increased reported uncertainty by approximately 0.31 to 0.54 scale points. Structure complexity led to the highest perceived difficulty (0.750***), again consistent with the dominance of simulation costs over storage costs as a source of subjective difficulty.

Effort. Importantly, increased complexity did not lead to a significant change in self-reported effort across any treatment arm. This invariance is consistent with a model in which subjects cannot identify in advance where the difficulty of a narrative lies—particularly for feedback loops, where the computational bottleneck may only become apparent after the subject has already attempted and failed to trace the causal logic. The absence of a compensatory effort response implies that the documented error increases are not offset by increased effort allocation, suggesting that complexity effects operate through cognitive architecture rather than motivational channels.

Table 3: OLS Results for Cognitive Uncertainty, Perceived Complexity, and Effort

	Cogn. Uncertainty		Perceived Comp.		Effort	
	(1)	(2)	(3)	(4)	(5)	(6)
State-1	0.320*** (0.057)	0.351*** (0.073)	0.486*** (0.075)	0.494*** (0.096)	-0.008 (0.040)	-0.028 (0.089)
State-2	0.541*** (0.062)	0.495*** (0.079)	0.526*** (0.079)	0.523*** (0.098)	-0.097 (0.129)	-0.131 (0.174)
Structure	0.511*** (0.044)	0.522*** (0.047)	0.711*** (0.057)	0.743*** (0.062)	-0.073 (0.064)	-0.103 (0.075)
Num.Obs.	1801	1511	1801	1511	1801	1511
Mean (Base. nar.)	2.71	2.71	4.39	4.39	9.23	9.23
Domain FE	Yes	Yes	Yes	Yes	Yes	Yes
Respondent FE	Yes	No	Yes	No	No	No
Demographic Controls	No	Yes	No	Yes	Yes	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.3 Cognitive Mechanisms: Representational vs. Computational Failure

We distinguish two distinct cognitive failure modes underlying our main results. *Representational failure* refers to the inability to correctly encode which variables and causal links exist in the narrative—a failure driven primarily by memory overload as the number of variables exceeds working memory capacity. *Computational failure* refers to the inability to correctly compute net causal effects from a (potentially correct) representation—a failure driven by the simulation cost required to resolve a feedback loop. The evidence we report below provide convergent evidence for the representational/computational distinction.

Isolating representational failure (Evidence I). We isolate the representational channel by focusing on error rates for Rulebook Question 1 (the $y \rightarrow a$ link), which is present in all treatment arms, never added or modified by any complexity treatment, and computationally trivial (requiring no traversal of the feedback loop). Any treatment-induced increase in Q1 error rates must therefore reflect pure representational failure—forgetting or misencoding a basic link—rather than computational difficulty.

Table 4 reports OLS estimates of complexity treatment effects on Q1 error rates, using the fictional and production domains only. State-2 significantly increases Q1 errors relative to baseline (0.037**), consistent with five-variable chains exceeding working memory capacity and leading to representational lapses. State-1 has a smaller and less significant effect. By contrast, the Structure treatment does not significantly increase Q1 error rates, consistent with the theoretical prediction that the feedback loop introduces a computational cost without increasing the representational burden (both Baseline and Structure contain the same three variables, and the $y \rightarrow a$ link is unaffected by the feedback).

Table 4: Isolating Representational Failure: Error Rate on Rulebook Q1 ($y \rightarrow a$ link)

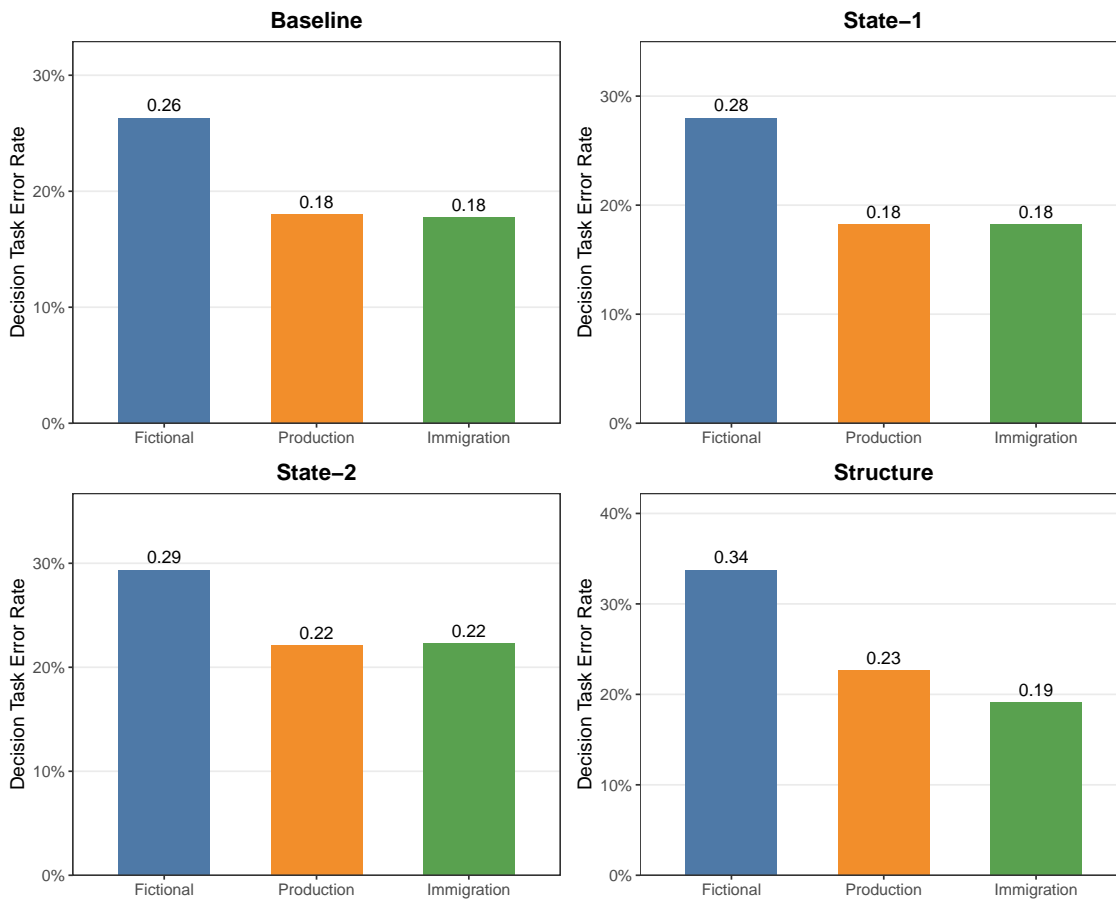
	Rulebook (a-b) Error Rate	
	(1)	(2)
State-1	0.016 (0.016)	0.025* (0.015)
State-2	0.030* (0.017)	0.032** (0.015)
Structure	0.015 (0.011)	0.017 (0.011)
Num.Obs.	1801	1511
Mean (Baseline)	0.08	0.08
Domain FE	Yes	Yes
Respondent FE	Yes	No
Demographic Controls	No	Yes

Notes: Sample restricted to fictional and production domains. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Cross-domain evidence for representational failure (Evidence II). A second, complementary identification strategy exploits cross-domain variation in representational difficulty. Within any given treatment arm, the causal structure is held constant across domains—only the variable labels change. Computational difficulty (which depends only on the causal structure) should therefore be invariant across domains within the same treatment arm. If representational failure is a primary driver of errors, error rates should vary systematically across domains within the same treatment arm, with the fictional domain exhibiting the *highest* error rates—because fictional variable labels (“aetheric energy flux,” “chrono-synaptic activation”) offer no real-world memory anchor and are harder to encode.

Figure 2 shows unconditional Decision Task error rates disaggregated by domain and complexity condition. Within all treatment arms, error rates are highest in the fictional domain and lower in production and immigration domains, confirming the predicted pattern. This within-treatment variation across domains is consistent with representational failure: subjects with better real-world anchors for variable concepts are better able to represent the causal structure in memory.

Figure 2: Unconditional Decision Task Error Rates by Domain and Complexity



Notes: Unconditional means of Decision Task error rates, disaggregated by domain and complexity treatment. Error rates are highest in the fictional domain within each treatment arm, consistent with representational difficulty being higher when variables have no real-world referent.

Isolating computational failure. We isolate the computational channel by conditioning on subjects who correctly encoded the $y \rightarrow a$ link (scored Rulebook Q1 correctly) as a proxy for representational accuracy. Among subjects who correctly represented at least this core link, residual errors in the Decision Task are more likely to arise from downstream computational mistakes—specifically, from the failure to resolve the feedback loop.

Table 5 presents Decision Task error rate regressions on complexity treatments, with and without a control for Q1 correctness. The key prediction is as follows: conditioning on Q1 correctness should reduce the State complexity coefficients (since these treatments primarily operate through representational failure), while leaving the Structure coefficient largely unchanged (since the feedback loop imposes a computational cost that persists even among accurate encoders). The results are consistent with this prediction: State-2 coefficients shrink

substantially when Q1 correctness is controlled for, while the Structure coefficient remains large and statistically significant.

Table 5: Isolating Computational Failure: Controlling for Representational Ability

	Decision Task Error Rate			
	(1)	(2)	(3)	(4)
State-1	0.010 (0.020)	0.008 (0.021)	0.003 (0.019)	0.003 (0.021)
State-2	0.036* (0.021)	0.037* (0.022)	0.027 (0.020)	0.030 (0.022)
Structure	0.061*** (0.016)	0.054*** (0.017)	0.057*** (0.015)	0.050*** (0.017)
Rulebook Q1 Correct			-0.290*** (0.029)	-0.216*** (0.033)
Num.Obs.	1801	1511	1801	1511
Mean (Baseline)	0.22	0.22	0.22	0.22
Domain Controls	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes

Notes: Sample restricted to fictional and production domains. Columns differ in whether Q1 correctness is included as a control. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.4 Immigration Domain

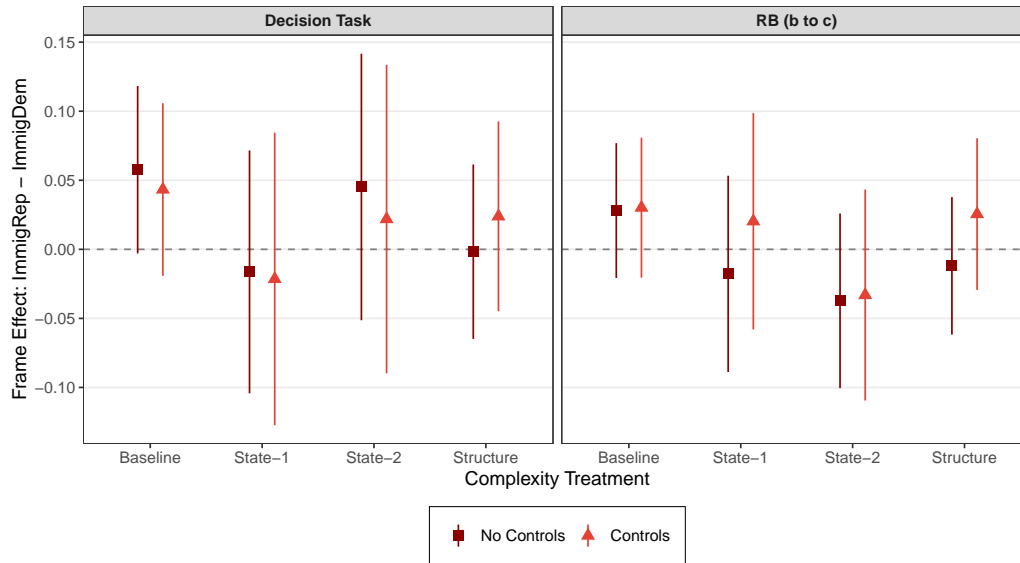
We examine the immigration domain separately because of its distinctive features: politically charged content, pre-existing strong prior beliefs among many subjects, and a randomized frame manipulation (Democrat- vs. Republican-leaning narrative). Complexity effects on error rates essentially vanish in the immigration domain—unlike in the fictional and production domains, complexity treatments do not significantly increase error rates among immigration narratives.

Figure 3 shows the difference in error rates between Democrat-leaning and Republican-leaning immigration frames, estimated separately by complexity treatment. At baseline, subjects assigned to the Democrat-leaning frame have somewhat different error rates than those assigned to the Republican-leaning frame. This difference varies with the complexity treatment, suggesting that narrative complexity interacts with political framing in a non-trivial way.

Figure 4 shows error rates disaggregated by respondent party affiliation (Democrat vs. Republican) and narrative frame (Democrat-leaning vs. Republican-leaning), for each complexity treatment. Two patterns stand out. First, political congruence—whether the subject’s party identity aligns with the narrative’s partisan slant—does not appear to interact with complexity in a meaningful way: the gap between congruent and incongruent subjects is broadly stable across the baseline, state, and structure treatment arms, suggesting that

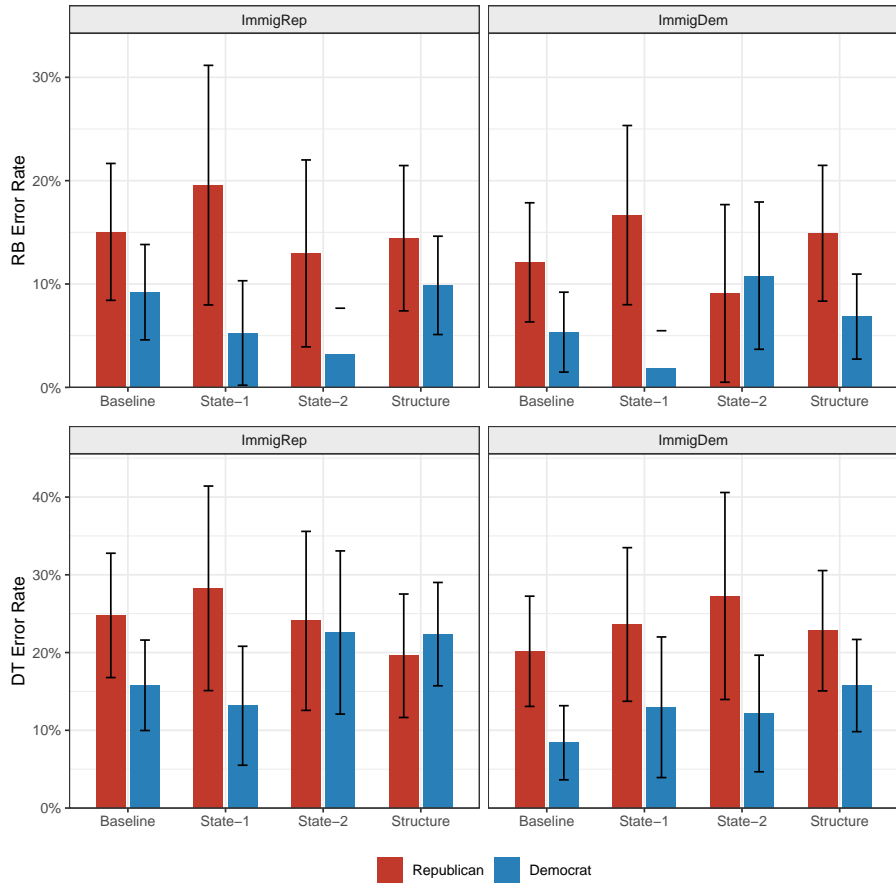
motivated reasoning operates largely independently of narrative structure in this setting. Second, Republicans exhibit systematically higher error rates than Democrats across frames and complexity conditions, and this gap persists after controlling for a rich set of individual-level characteristics including gender, education, income, and IQ score, indicating that party affiliation captures something beyond observable cognitive or socioeconomic differences.

Figure 3: Immigration Frame Effect by Complexity Treatment



Notes: The figure shows the difference in error rates between the Democrat-leaning and Republican-leaning immigration frames, estimated separately by complexity treatment.

Figure 4: Immigration Error Rates by Treatment and Party Affiliation



Notes: Error rates broken down by respondent party affiliation (Democrat vs. Republican) and complexity treatment, shown separately for the Democrat-leaning and Republican-leaning immigration frames.

4.5 Note-Taking Behavior

Taking notes during the 2-minute briefing period is associated with lower error rates on both the Rulebook and Decision Taskç Table 6 and 7 report estimates of the association between note-taking (binary: whether the subject recorded any text in the notes field), character count, word count, and error rates.

Table 6: Note-taking and decision task performance

	DT Error Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Took Notes	-0.019 (0.043)	-0.080** (0.036)				
Char. Count			-0.001 (0.001)	0.000 (0.001)		
Word Count					-0.001 (0.003)	-0.001 (0.002)
Num.Obs.	1801	1511	1745	1464	1745	1464
Mean (Base. nar.)	0.21	0.21	0.20	0.20	0.20	0.20
Sample	Full	Full	Note-takers	Note-takers	Note-takers	Note-takers
Domain FE	Yes	Yes	Yes	Yes	Yes	Yes
Complexity FE	Yes	Yes	Yes	Yes	Yes	Yes
Respondent FE	Yes	No	Yes	No	Yes	No
Controls	No	Yes	No	Yes	No	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Note-taking and rulebook performance

	RB (b -> c) Error Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Took Notes	0.016 (0.043)	-0.070** (0.030)				
Char. Count			-0.001 (0.001)	-0.002*** (0.001)		
Word Count					-0.004 (0.003)	-0.006*** (0.002)
Num.Obs.	1801	1511	1745	1464	1745	1464
Mean (Base. nar.)	0.11	0.11	0.11	0.11	0.11	0.11
Sample	Full	Full	Note-takers	Note-takers	Note-takers	Note-takers
Domain FE	Yes	Yes	Yes	Yes	Yes	Yes
Complexity FE	Yes	Yes	Yes	Yes	Yes	Yes
Respondent FE	Yes	No	Yes	No	Yes	No
Controls	No	Yes	No	Yes	No	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.6 Awareness Gap

We construct an *awareness-of-complexity* measure as the difference between each subject's standardized perceived complexity (PC) and their standardized rulebook error rate, following [Agranov et al. \(2025\)](#). A positive awareness gap indicates that the subject perceives the task

as harder than their actual error rate would suggest; a negative gap indicates overconfidence. Table 8 reports OLS estimates of complexity treatment effects on the awareness gap.

Table 8: Awareness of Complexity

	Awareness Gap		Awareness Gap	
	(1)	(2)	(3)	(4)
State-1	0.117*** (0.044)	0.104** (0.048)	0.083*** (0.029)	0.071** (0.030)
State-2	0.114** (0.045)	0.092* (0.051)	0.049* (0.029)	0.095*** (0.031)
Structure	0.086** (0.034)	0.086** (0.037)	0.060*** (0.022)	0.076*** (0.023)
Fictional	0.067** (0.034)	0.063* (0.037)	0.103*** (0.022)	0.105*** (0.024)
Immigration	-0.035 (0.032)	-0.039 (0.035)	-0.050** (0.021)	-0.046** (0.022)
Male		-0.213*** (0.051)		-0.015 (0.028)
Age		0.010*** (0.002)		0.004*** (0.001)
IQ Score		0.017 (0.019)		-0.062*** (0.010)
Num.Obs.	1801	1511	1801	1511
Mean (Base. nar.)	-0.07	-0.07	1.02	1.02
Domain FE	Yes	Yes	Yes	Yes
Respondent FE	Yes	No	Yes	No
Demographic Controls	No	Yes	No	Yes

Notes: Awareness of complexity is calculated as the difference between standardized perceived complexity and standardized RB error rate. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5 Conclusion

This paper provides the first experimental decomposition of narrative complexity into theoretically grounded structural dimensions and measures their causal effects on comprehension and decision-making. We study how the structural properties of causal narratives affect citizens' ability to understand and act upon policy-relevant information—a question that bears directly on the quality of political discourse and the demand for good policy.

Our central finding is that not all complexity is alike. Two dimensions of structural complexity that we independently manipulate—the number of variables in a linear causal chain (state complexity) and the presence of a feedback loop (structure complexity)—have qualitatively different cognitive effects. State complexity, which progressively exceeds working memory capacity as the chain lengthens, generates modest effects that are consistent with *representational failure*: when subjects must track more variables than their memory can

reliably hold, they begin to misencode the causal structure. Structure complexity, which introduces mutual causation through a feedback loop, generates substantially larger effects that are consistent with *computational failure*: even subjects who correctly encode the basic causal links struggle to compute net effects through a system where causes and consequences are co-determined.

These failures are economically costly. Both error measures—Rulebook comprehension errors and Decision Task errors—rise significantly when structural complexity increases. The absence of a compensatory effort response implies that subjects are not aware of the full extent of their comprehension difficulties. Subjects experience the cognitive burden of structural complexity but do not fully perceive it: their subjective complexity ratings increase, but not by as much as their error rates.

A particularly striking finding is the complete attenuation of complexity effects in the immigration domain. In the fictional and production settings, where subjects have no strong pre-existing views about the causal relationships described in the narratives, complexity systematically degrades performance. In the immigration setting, however, prior political beliefs appear to dominate causal reasoning: regardless of the structural complexity of the narrative, subjects’ error rates are primarily predicted by their political identity and the narrative’s partisan frame rather than by whether the narrative contains a feedback loop. This finding is consistent with a model in which strong prior beliefs function as a “cognitive shortcut” that bypasses effortful causal reasoning—a shortcut that is activated under both simple and complex narratives in politically charged settings.

Taken together, our results carry two broad implications. First, from a policy design perspective, they suggest that the structural simplification of policy arguments—reducing feedback complexity rather than merely shortening causal chains may be the most powerful lever for improving public comprehension of policy tradeoffs outside of politically loaded issues. Feedback loops are endemic to accurate representations of general equilibrium effects, which means that factually complete policy communication is inevitably harder to process than simplified, linear narratives that strip out the recursive consequences of policy interventions. Understanding this trade-off is essential for designing public communication that is both accurate and accessible. Second, from a research perspective, our results provide a structural micro-foundation for the empirically documented demand for policies with adverse equilibrium consequences (Dal Bó et al., 2018; Nunnari et al., 2024): citizens who cannot process feedback loops in even a simple laboratory narrative would, by extension, systematically underestimate the equilibrium consequences of real-world policy interventions communicated

through more complex real-world narratives.

References

- Agranov, M., Schotter, A., and Trevino, I. (2025). “Complex for whom? an experimental approach to subjective complexity”. In: *Available at agranov.caltech.edu*.
- Ambuehl, S., Bhui, R., and Thysen, H. C. (2026). “Mental models of causal structure in economics and psychology”. In: *arXiv preprint arXiv:2603.29070*.
- Ambuehl, S. and Thysen, H. C. (2025). “Choice with Competing Models: An Experimental Study”. In.
- Andre, P., Haaland, I., Roth, C., Wiederholt, M., and Wohlfart, J. (2024). *Narratives about the Macroeconomy*. Tech. rep. SAFE Working Paper.
- Andre, P., Pizzinelli, C., Roth, C., and Wohlfart, J. (2022). “Subjective models of the macroeconomy: Evidence from experts and representative samples”. In: *The Review of Economic Studies* 89.6, pp. 2958–2991.
- Arrieta, G. and Nielsen, K. (2024). *Procedural decision-making in the face of complexity*. Tech. rep. Working Paper.
- Ba, C., Bohren, J. A., and Imas, A. (2024). “Over-and underreaction to information”. In: *Available at SSRN 4274617*.
- (2025). “The role of representational and computational complexity in belief formation”. In: *AEA Papers and Proceedings*. Vol. 115. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 631–636.
- Banovetz, J. and Oprea, R. (2023). “Complexity and procedural choice”. In: *American Economic Journal: Microeconomics* 15.2, pp. 384–413.
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., and Husain, M. (2011). “Temporal dynamics of encoding, storage, and reallocation of visual working memory”. In: *Journal of vision* 11.10, pp. 6–6.
- Benjamin, D. J. (2019). “Errors in probabilistic reasoning and judgment biases”. In: *Handbook of Behavioral Economics: Applications and Foundations 1 2*, pp. 69–186.
- Blesse, S., Gruendler, K., Heil, P., and Hermes, H. (2025). “The demand for economic narratives”. In.
- Bohren, J. A., Hascher, J., Imas, A., Ungeheuer, M., and Weber, M. (2024). “Mental Models of Information: The Role of Attention and Memory”. In: *Becker Friedman Institute for Economics Working Paper* 19, p. 93.

- Charles, C. and Kendall, C. (2025). “Causal narratives”. In: *Available at SSRN 4669371*.
- Clippel, G. de, Moscariello, P., Ortoleva, P., and Rozen, K. (2024). *Caution in the face of complexity*. Tech. rep. Mimeo.
- Dal Bó, E., Dal Bó, P., and Eyster, E. (2018). “The demand for bad policy when voters underappreciate equilibrium effects”. In: *The Review of Economic Studies* 85.2, pp. 964–998.
- Eliasz, K. and Spiegler, R. (2020). “A model of competing narratives”. In: *American Economic Review* 110.12, pp. 3786–3816.
- Fréchette, G. R., Vespa, E., and Yuksel, S. (2024). “Extracting models from data sets: An experiment”. In: *Available at SSRN 5026751*.
- Gabaix, X. and Graeber, T. (2024). *The complexity of economic decisions*. Tech. rep. National Bureau of Economic Research.
- Halford, G. S., Baker, R., McCredden, J. E., and Bain, J. D. (2005). “How many variables can humans process?” In: *Psychological science* 16.1, pp. 70–76.
- Kalaycı, K. and Serra-Garcia, M. (2016). “Complexity and biases”. In: *Experimental Economics* 19.1, pp. 31–50.
- Kastens, K. A. and Shipley, T. F. (2023). “How does the human mind think and learn about feedback loops”. In: *Proceedings of the International Conference on Systems Dynamics, Chicago, IL, Systems Dynamics Society*.
- Kendall, C. and Oprea, R. (2024). “On the complexity of forming mental models”. In: *Quantitative Economics* 15.1, pp. 175–211.
- Kendall, C. W. and Charles, C. (2022). *Causal narratives*. Tech. rep. National Bureau of Economic Research.
- Levy, G., Razin, R., and Young, A. (2022). “Misspecified politics and the recurrence of populism”. In: *American Economic Review* 112.3, pp. 928–962.
- Loewenstein, G. and Wojtowicz, Z. (2025). “The economics of attention”. In: *Journal of Economic Literature* 63.3, pp. 1038–1089.
- Lombrozo, T. and Vasilyeva, N. (2017). “22 Causal Explanation”. In: *The Oxford handbook of causal reasoning*, p. 415.
- Luck, S. J. and Vogel, E. K. (1997). “The capacity of visual working memory for features and conjunctions”. In: *Nature* 390.6657, pp. 279–281.
- MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

- Meder, B. and Mayrhofer, R. (2017). “23 Diagnostic Reasoning”. In: *The Oxford handbook of causal reasoning*, p. 433.
- Nam, A., Hughes, C., Icard, T., and Gerstenberg, T. (2023). “Show and tell: Learning causal structures from observations and explanations”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 45. 45.
- Nunnari, S., Proto, E., and Rustichini, A. (2024). *Cognitive abilities and the demand for bad policy*. IZA-Institute of Labor Economics.
- Oberauer, K., Farrell, S., Jarrold, C., and Lewandowsky, S. (2016). “What limits working memory capacity?” In: *Psychological bulletin* 142.7, p. 758.
- Oprea, R. (2020). “What makes a rule complex?” In: *American economic review* 110.12, pp. 3913–3951.
- (2024). “Decisions under risk are decisions under complexity”. In: *American Economic Review* 114.12, pp. 3789–3811.
- Rottman, B. M. and Hastie, R. (2014). “Reasoning about causal relationships: Inferences on causal networks.” In: *Psychological bulletin* 140.1, p. 109.
- Shiller, R. J. (2020). “Narrative economics: How stories go viral and drive major economic events”. In.
- Sloman, S., Barbey, A. K., and Hotaling, J. M. (2009). “A causal model theory of the meaning of cause, enable, and prevent”. In: *Cognitive Science* 33.1, pp. 21–50.

- A Online Appendix: Additional Tables
- B Online Appendix: Additional Figures
- C Online Appendix: Survey Instructions