Information Overload, Cognitive Capacity, and Customer Engagement:
Evidence from Retail Messaging Data

1 Introduction and Major Research Question

Digital retail platforms rely heavily on repeated messaging—emails, push notifications, and in-app notifications—to stimulate customer engagement. However, customers operate in an environment of limited attention: each incoming commercial message competes with other personal messages, market-wide advertising, news events, and daily cognitive fatigue. As a result, opening a message is not automatic. It is a decision under uncertainty, where customers weigh the expected value of opening against the cognitive cost of processing it.

This project studies how different sources of information pressure—including a firm's own messaging frequency, competitor-side market pressure, contextual timing (weekday, hour, holidays), and broader information shocks—affect customers' cognitive capacity, fatigue, and ultimately message-opening behavior.

The central research question is:

> **How do multiple information-environment factors shape customer cognitive capacity and thereby influence message-opening behavior in retail settings?**

A secondary practical question is:

> **How do these effects differ across customer segments, and how should firms adjust messaging strategies accordingly?**

Understanding this helps firms avoid oversupplying commercial information, which can lead to customer fatigue, disengagement, and reduced campaign ROI.

2 Conceptual Framework: Decision Under Uncertainty

We model customer i at time t receiving a message and choosing action $a_{it} \in \{0, 1\}$ where $a_{it} = 1$ means opening the message. Following standard decision-under-uncertainty logic, the customer evaluates:

$$U_{it}(1) = E_{it}[V_{it}] - C_{it}, \qquad U_{it}(0) = 0.$$

Thus,

$$a_{it} = 1 \iff E_{it}[V_{it}] - C_{it} \geq 0.$$

Decomposition of Expected Value

We assume the customer's expected value of opening can be written as:

$$E_{it}[V_{it}] = \theta_i \cdot q_{it}.$$

The components have natural interpretations in our dataset:

- θ_i**: latent preference or general inclination to engage.** This is a stable parameter capturing long-run affinity for offers. *Dataset proxies:*
  - lifetime open rate,
  - past purchase history / click frequency,
  - historical engagement segment (loyal, occasional, dormant).
- q_{it}**: message-specific relevance signal.** This varies across messages and contexts. *Dataset proxies:*
  - campaign-level historical open rates,
  - subject-line features (discount, personalization),
  - customer's short-run open signals (1-week and 1-month open rates),
  - contextual relevance: time of day, weekday, holiday indicators.

Thus, decomposition clarifies that customers open messages when:

$$\text{long-run interest} \times \text{short-run relevance} \geq \text{cognitive cost}.$$

Decomposition of Cognitive Cost

We model attention cost as:

$$C_{it} = \alpha_i + \beta F_{it}.$$

- α_i: **baseline cognitive cost** Captures stable differences in attention capacity or habitual responsiveness. *Dataset proxies:*
  - segment indicators (loyal vs dormant),
  - long-term behavioral variability.
- F_{it}: **dynamic fatigue state** driven by information pressure. We specify:

$$F_{i,t} = \rho F_{i,t-1} + \gamma \,\text{MsgsSent}_{i,t-1} + \delta \, M_t + \varepsilon_{it}.$$

  Dataset proxies:
  - recent messages sent to customer (1-day, 7-day, 30-day counts),
  - recency since last message,
  - market-wide messaging averages (1-week, 1-month),
  - contextual load factors: weekday, hour, news/event periods.

These factors collectively shift the cognitive cost of evaluating a new message.

Agents and Their Interaction Structure

The economic environment consists of:

- **Customers (agents)** choosing whether to open each incoming message.
- **The firm** deciding messaging frequency and content.
- **The market** generating external messaging pressure (competitors, global advertising intensity).

Interaction arises because:

- customers face simultaneous information pressure from the firm and from the broader market;
- the firm's decision creates cognitive load that affects future customer actions;
- customers' heterogeneous θ_i and α_i produce different responses even under identical pressure.

Customers maximize experienced utility (expected relevance minus cognitive cost). The firm aims to maximize engagement (open probability), subject to not exhausting customers' attention. The market is an exogenous source of information shocks.

This environment produces a natural tension between engagement and information overload, justifying the empirical analysis.

3 Testable Predictions

From the framework, we obtain three primary predictions:

1. **Higher information pressure reduces engagement.** Increased messaging (own or market-wide) raises F_{it} and therefore C_{it}, decreasing open probability.

2. **Heterogeneous fatigue effects across customer segments.** Loyal/high-engagement customers have lower α_i and lower fatigue sensitivity. Dormant customers show strong fatigue effects even at low messaging frequency.

3. **Contextual timing modifies cognitive cost.** Weekends vs weekdays, work hours vs leisure hours, and holidays shift C_{it}, creating predictable timing effects.

These align with rational behavior under scarcity of attention.

4 Data Plan

We use fully engineered internal datasets. Outcome:

- binary indicator: message opened.

Explanatory variables:

Own-Firm Pressure / Fatigue Proxies

- message counts: last 1 day, 7 days, 30 days,
- recency since last message,
- volatility of messaging exposure,
- short-term / long-term fatigue ratios.

Market-Wide Pressure

- market-average messaging in last 1 week and 1 month (simulated or aggregated),
- external news-cycle variables (if available).

Contextual Cognitive Load Shifters

- day-of-week,
- time-of-day,
- holiday indicators or special event windows.

Expected-Value Proxies

- customer short-run open rates (1 week, 1 month),
- lifetime engagement,
- campaign-level smoothed open rates,
- subject-line features (discount, personalization).

Segmentation Variables

Three segments:

1. loyal (high engagement),
2. occasional,
3. dormant (low engagement).

5 Empirical Strategy

1. Descriptive Exploration

- plots of open rates vs fatigue measures,
- comparison of fatigue curves across segments,
- timing effects visualized across days and hours.

2. Reduced-Form Econometric Model

We estimate:

$$\Pr(\text{open}_{it} = 1) = \Lambda\big(\beta_1 \text{Fatigue}_{it} + \beta_2 \text{ExpectedValue}_{it} + \beta_3 (\text{Fatigue}_{it} \times \text{Segment}_i) + X_{it}\gamma\big)$$

where X_{it} includes contextual controls and campaign fixed effects.

3. Heterogeneous Effects and Counterfactuals

- estimation by customer segment,
- simulation of predicted open rates under reduced messaging (10%, 20%, 30% decreases),
- recommendations for optimal cadence by segment.

6 Expected Outputs

Figures

- fatigue curves by segment,
- heatmap of predicted open probability (fatigue × expected value),
- contextual timing effects.

Tables

- regression results with marginal effects,
- heterogeneous segment effects,
- counterfactual simulation results.

Contributions

- show how multiple information-pressure sources jointly determine cognitive capacity,
- highlight strong heterogeneity across segments,
- produce actionable guidance for campaign strategy design.

7 Feasibility

We already possess:

- engineered fatigue and expected-value proxies,
- market pressure variables,
- segmentation,
- experience with logit/probit models,
- clean panel data structure.